GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 26, 2005, 16:30:06 ; Search time 42 Seconds
(without alignments)
447.894 Million cell updates/sec

Title: US-10-774-076-1
Perfect score: 1308
Sequence: 1 MRAPLLPPAPVVLSLLILGS.....GEAEERKKLRQENGNVHAIA 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1308	100.0	252	2	US-07-885-089B-7	Sequence 7, Appli
2	1308	100.0	252	4	US-09-976-594-582	Sequence 582, App
3	1308	100.0	252	4	US-09-919-039-228	Sequence 228, App
4	1308	100.0	322	4	US-09-949-016-11381	Sequence 11381, A
5	1177	90.0	226	2	US-07-885-089B-35	Sequence 35, Appl
6	1015.5	77.6	192	1	US-08-208-008C-9	Sequence 9, Appli
7	1004.5	76.8	247	2	US-07-885-089B-2	Sequence 2, Appli
8	1004.5	76.8	247	2	US-07-885-089B-8	Sequence 8, Appli
9	914	69.9	248	2	US-07-885-089B-6	Sequence 6, Appli
10	914	69.9	248	2	US-07-885-089B-10	Sequence 10, Appl
11	874.5	66.9	243	2	US-07-885-089B-4	Sequence 4, Appli
12	874.5	66.9	243	2	US-07-885-089B-9	Sequence 9, Appli
13	874.5	66.9	243	6	5202428-2	Patent No. 5202428
14	874.5	66.9	243	6	5202428-2	Patent No. 5202428
15	855	65.4	158	2	US-07-885-089B-34	Sequence 34, Appl
16	789	60.3	152	2	US-07-885-089B-36	Sequence 36, Appl
17	762	58.3	146	2	US-07-885-089B-37	Sequence 37, Appl
18	700.5	53.6	137	6	5202428-9	Patent No. 5202428
19	700.5	53.6	137	6	5202428-9	Patent No. 5202428
20	534.5	40.9	147	6	5202428-8	Patent No. 5202428
21	534.5	40.9	147	6	5202428-8	Patent No. 5202428
22	501	38.3	91	3	US-09-077-977A-1	Sequence 1, Appli
23	467	35.7	84	6	5202428-10	Patent No. 5202428
24	467	35.7	84	6	5202428-10	Patent No. 5202428
25	459	35.1	83	2	US-07-885-089B-16	Sequence 16, Appl
26	454	34.7	87	1	US-07-847-743B-17	Sequence 17, Appl
27	454	34.7	87	1	US-08-456-201-17	Sequence 17, Appl
28	454	34.7	87	2	US-08-330-161-15	Sequence 15, Appl
29	454	34.7	87	2	US-08-456-241-17	Sequence 17, Appl
30	454	34.7	87	2	US-08-440-401-15	Sequence 15, Appl
31	454	34.7	87	2	US-08-419-878B-15	Sequence 15, Appl
32	454	34.7	87	3	US-09-173-480-15	Sequence 15, Appl
33	454	34.7	87	5	PCT-US92-04295A-17	Sequence 17, Appl
34	442	33.8	83	2	US-07-885-089B-18	Sequence 18, Appl
35	440	33.6	79	2	US-07-885-089B-33	Sequence 33, Appl
36	433	33.1	85	2	US-07-885-089B-11	Sequence 11, Appl
37	356	27.2	78	2	US-07-885-089B-32	Sequence 32, Appl
38	355	27.1	78	2	US-07-885-089B-30	Sequence 30, Appl
39	347.5	26.6	79	2	US-07-885-089B-31	Sequence 31, Appl
40	320	24.5	87	1	US-07-847-743B-18	Sequence 18, Appl
41	320	24.5	87	1	US-08-456-201-18	Sequence 18, Appl
42	320	24.5	87	2	US-08-330-161-16	Sequence 16, Appl
43	320	24.5	87	2	US-08-456-241-18	Sequence 18, Appl
44	320	24.5	87	2	US-08-440-401-16	Sequence 16, Appl
45	320	24.5	87	2	US-08-419-878B-16	Sequence 16, Appl

ALIGNMENTS

```
RESULT 1
US-07-885-089B-7
; Sequence 7, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  7:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  252 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-885-089B-7

  Query Match           100.0%;  Score 1308;  DB 2;  Length 252;
  Best Local Similarity  100.0%;  Pred. No. 8.2e-122;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
```

```
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252

Db        241 LRQENGNVHAIA 252
```

RESULT 2
US-09-976-594-582
; Sequence 582, Application US/09976594
; Patent No. 6673549
; GENERAL INFORMATION:
; APPLICANT: Furness, Michael
; APPLICANT: Buchbinder, Jenny
; TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
; FILE REFERENCE: PA-0041 US
; CURRENT APPLICATION NUMBER: US/09/976,594
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: 60/240,409
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 1143
; SOFTWARE: PERL Program
; SEQ ID NO 582
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6673549 3478236CD1
US-09-976-594-582

Query Match 100.0%; Score 1308; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 8.2e-122;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252

Db        241 LRQENGNVHAIA 252
```

RESULT 3
US-09-919-039-228
, Sequence 228, Application US/09919039
, Patent No. 6727066
; GENERAL INFORMATION:
; APPLICANT: Kaser, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919,039
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 228
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6727066 3478236CD1
US-09-919-039-228

Query Match 100.0%; Score 1308; DB 4; Length 252;
Best Local Similarity 100.0%; Pred. No. 8.2e-122;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252

Db        241 LRQENGNVHAIA 252
```

RESULT 4
US-09-949-016-11381
; Sequence 11381, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11381
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-11381

Query Match 100.0%; Score 1308; DB 4; Length 322;
Best Local Similarity 100.0%; Pred. No. 1.2e-121;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         71 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 130

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        131 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 190

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        191 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 250

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        251 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 310

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        311 LRQENGNVHAIA 322
```

RESULT 5
US-07-885-089B-35
; Sequence 35, Application US/07885089B
; Patent No. 5830995
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: McDonald, Vicki L.
; APPLICANT: Bradley, James G.
; APPLICANT: Plowman, Gregory D.
; TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF
; TITLE OF INVENTION: HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,089B
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 5624-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 35:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 226 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-885-089B-35

Query Match 90.0%; Score 1177; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 7.3e-109;
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy         27 LDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEE 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEE 60

Qy         87 YDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPC 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPC 120

Qy        147 NAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAA 206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 NAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAA 180

Qy        207 FMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVHAIA 252
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 FMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVHAIA 226
```

RESULT 6
US-08-208-008C-9
; Sequence 9, Application US/08208008C
; Patent No. 5633147
; GENERAL INFORMATION:
; APPLICANT: MEISSNER, ET AL.
; TITLE OF INVENTION: Transforming Growth Factor
; TITLE OF INVENTION: Alpha - H1
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/208,008C
; FILING DATE: March 8, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: MULLINS, J.G.
; REGISTRATION NUMBER: 33,073
; REFERENCE/DOCKET NUMBER: 325800-98
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 192 AMINO ACIDS
; TYPE: AMINO ACID
; STRANDEDNESS:
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
US-08-208-008C-9

Query Match 77.6%; Score 1015.5; DB 1; Length 192;
Best Local Similarity 99.5%; Pred. No. 6.5e-93;
Matches 192; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLN-TYSGKREPFSGDHSADGFEVTSRSEMSSG 59

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         60 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 119

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        120 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 179

Qy        181 CGEKSMKTHSMID 193
              |||||||||||||
Db        180 CGEKSMKTHSMID 192
```

RESULT 7
US-07-885-089B-2
; Sequence 2, Application US/07885089B
; Patent No. 5830995
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: McDonald, Vicki L.
; APPLICANT: Bradley, James G.
; APPLICANT: Plowman, Gregory D.
; TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF
; TITLE OF INVENTION: HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,089B
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 5624-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-885-089B-2

Query Match 76.8%; Score 1004.5; DB 2; Length 247;
Best Local Similarity 77.5%; Pred. No. 1.2e-91;
Matches 196; Conservative 19; Mismatches 31; Indels 7; Gaps 2;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              |||||||||||||||||| || || ||||:||||||| |||||||||| |||||||:||
Db          1 MRAPLLPPAPVVLSLLIFGSAHYTAGLDVNDTYSGKGEPFSGDHSADRFEVTSRSEISSA 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||  |  |:      ||  ||||:||||||||| |||||||||||||||| :|||||| |
Db         61 SETPPGGEL------SSVIDYDYAEEYDNEPQISGYIVDDSVRVEQVVKPKKNKTESEKT 114

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              |||||||||||||||||||||| |: |||||||||:| ::| || |:|:|  ||||||
Db        115 SDKPKRKKKGGKNGKNRRNRKKKNLCDTEFQNFCIHGKCTFLEQLETVSCQCYPEYFGER 174

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAV-ITVQLRRQYVRKYEGEAEERK 239
              |||||||| ||:|| ||||||||||:||:  ||:|| ||: |||: :| ||| |||||
Db        175 CGEKSMKTQSMVDSDLSKIALAAIAAFVSAMTFTAIAVFITILLRRRCLRGYEGVAEERK 234

Qy        240 KLRQENGNVHAIA 252
              |||||||| ||:|
Db        235 KLRQENGNAHAVA 247
```

RESULT 8
US-07-885-089B-8
; Sequence 8, Application US/07885089B
; Patent No. 5830995
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: McDonald, Vicki L.
; APPLICANT: Bradley, James G.
; APPLICANT: Plowman, Gregory D.
; TITLE OF INVENTION: AMPHIREGULINS: A FAMILY OF
; TITLE OF INVENTION: HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/885,089B
; FILING DATE: 18-MAY-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 5624-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 247 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
US-07-885-089B-8

Query Match 76.8%; Score 1004.5; DB 2; Length 247;
Best Local Similarity 77.5%; Pred. No. 1.2e-91;
Matches 196; Conservative 19; Mismatches 31; Indels 7; Gaps 2;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              |||||||||||||||||| || || ||||:||||||| |||||||||| |||||||:||
Db          1 MRAPLLPPAPVVLSLLIFGSAHYTAGLDVNDTYSGKGEPFSGDHSADRFEVTSRSEISSA 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||  |  |:      ||  ||||:||||||||| |||||||||||||||| :|||||| |
Db         61 SETPPGGEL------SSVIDYDYAEEYDNEPQISGYIVDDSVRVEQVVKPKKNKTESEKT 114

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              |||||||||||||||||||||| |: |||||||||:| ::| || |:|:|  ||||||
Db        115 SDKPKRKKKGGKNGKNRRNRKKKNLCDTEFQNFCIHGKCTFLEQLETVSCQCYPEYFGER 174

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAV-ITVQLRRQYVRKYEGEAEERK 239
              |||||||| ||:|| ||||||||||:||:  ||:|| ||: |||: :| ||| |||||
Db        175 CGEKSMKTQSMVDSDLSKIALAAIAAFVSAMTFTAIAVFITILLRRRCLRGYEGVAEERK 234

Qy        240 KLRQENGNVHAIA 252
              |||||||| ||:|
Db        235 KLRQENGNAHAVA 247
```

```
RESULT 9
US-07-885-089B-6
; Sequence 6, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  248 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-885-089B-6

  Query Match             69.9%;  Score 914;  DB 2;  Length 248;
  Best Local Similarity   69.0%;  Pred. No. 1.1e-82;
  Matches  176;  Conservative   28;  Mismatches   41;  Indels   10;  Gaps    3;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPPSGDHSADGFEVTSRSEMSSG 60
              || |||| |  || ||:|||||||| |:|||  ||| |  |||||| |       |:| |
Db          1 MRTPLLPLARSVLLLLVLGSGHYAAALELNDPSSGKGESLSGDHSAGGL------ELSVG 54

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
               |:| :||||| || |:| |||||||||||||| |||:|||||||||:|| :|||| | :
Db         55 REVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIIDDSVRVEQVIKPKKNKTEGEKS 113

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ::|||||||||||||: |||:|||||| |:||||||||||:|||:|| ||| | |:|||||
Db        114 TEKPKRKKKGGKNGEGRRNKKKKNPCTAKFQNFCIHGECRYIENLEVVTCNCHQDYFGER 173

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVA---VITVQLRRQYVRKYEGEAEE 237
              ||||||||||  |  |||||: |:  |:||:|| |:    |||| | ::| |:|||| ||
Db        174 CGEKSMKTHSEDDKDLSKIAVVAVTIFVSAIILAAIGIGIVITVHLWKRYFREYEGETEE 233

Qy        238 RKKLRQENGNVHAIA 252
              |::|||||| |||||
Db        234 RRRLRQENGTVHAIA 248


RESULT 10
US-07-885-089B-10
; Sequence 10, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  248 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-885-089B-10

  Query Match             69.9%;  Score 914;  DB 2;  Length 248;
  Best Local Similarity   69.0%;  Pred. No. 1.1e-82;
  Matches  176;  Conservative   28;  Mismatches   41;  Indels   10;  Gaps    3;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              || |||| |  || ||:|||||||| |:|||  ||| |  |||||| |       |:| |
Db          1 MRTPLLPLARSVLLLLVLGSGHYAAALELNDPSSGKGESLSGDHSAGGL------ELSVG 54

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
               |:| :||||| || |:| |||||||||||||| |||:|||||||||:|| :|||| | :
Db         55 REVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIIDDSVRVEQVIKPKKNKTEGEKS 113

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ::|||||||||||||: |||:|||||| |:||||||||||:|||:|| ||| | |:|||||
Db        114 TEKPKRKKKGGKNGEGRRNKKKKNPCTAKFQNFCIHGECRYIENLEVVTCNCHQDYFGER 173

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVA---VITVQLRRQYVRKYEGEAEE 237
              ||||||||||  |  |||||: |:  |:||:|| |:    |||| | ::| |:|||| ||
Db        174 CGEKSMKTHSEDDKDLSKIAVVAVTIFVSAIILAAIGIGIVITVHLWKRYFREYEGETEE 233

Qy        238 RKKLRQENGNVHAIA 252
              |::|||||| |||||
Db        234 RRRLRQENGTVHAIA 248


RESULT 11
US-07-885-089B-4
; Sequence 4, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
```

```
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-885-089B-4

  Query Match           66.9%;  Score 874.5;  DB 2;  Length 243;
  Best Local Similarity  69.7%;  Pred. No. 9.3e-79;
  Matches  177;  Conservative   24;  Mismatches   40;  Indels   13;  Gaps    5;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||: | |  | |||||:||||||||||:|| | ||| || |||||| |  |
Db          1 MRTPSLSLALSVLSLLVLGSGHYAAGLELNGTSSGKGEPSSGDHSAGGLVV--------- 51

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||:| :||||| || |:| ||||||||||||| ||||||||||||:|| :|||| | :
Db         52 SEVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIVDDSVRVEQVIKPKENKTEGEKS 110

Qy        121 SDKPKRKKKGGKNGKNRRNR-KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGE 179
              |:|||||||||| || |||| |||||| |:|||||||||:|||:|| ||| | |:||||
Db        111 SEKPKRKKKGGKGGKGRRNRKKKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGE 170

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVIT-VQLRRQYVRKYEGEAEER 238
              |||||:|||    || |||||||||  |:||| : |: :|| | ||::: |:|| |||||
Db        171 RCGEKTMKTQKKDDSDLSKIALAAIIVFVSAVSVAAIGIITAVLLRKRFFREYE-EAEER 229

Qy        239 KKLRQENGNVHAIA 252
              ::|||||  ||||
Db        230 RRLRQENGTAHAIA 243


RESULT 12
US-07-885-089B-9
; Sequence 9, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  243 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-885-089B-9

  Query Match           66.9%;  Score 874.5;  DB 2;  Length 243;
  Best Local Similarity  69.7%;  Pred. No. 9.3e-79;
  Matches  177;  Conservative   24;  Mismatches   40;  Indels   13;  Gaps    5;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              || | |  | |||||:||||||||||:|| | ||| || |||||| |  |
Db          1 MRTPSLSLALSVLSLLVLGSGHYAAGLELNGTSSGKGEPSSGDHSAGGLVV--------- 51

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||:| :||||| || |:| ||||||||||||| ||||||||||||:|| :|||| | :
Db         52 SEVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIVDDSVRVEQVIKPKENKTEGEKS 110

Qy        121 SDKPKRKKKGGKNGKNRRNR-KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGE 179
              |:|||||||||| || |||| |||||| |:|||||||||:|||:|| ||| | |:||||
Db        111 SEKPKRKKKGGKGGKGRRNRKKKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGE 170

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVIT-VQLRRQYVRKYEGEAEER 238
              |||||:|||    || |||||||||  |:||| : |: :|| | ||::: |:|| |||||
Db        171 RCGEKTMKTQKKDDSDLSKIALAAIIVFVSAVSVAAIGIITAVLLRKRFFREYE-EAEER 229

Qy        239 KKLRQENGNVHAIA 252
              ::|||||  ||||
Db        230 RRLRQENGTAHAIA 243


RESULT 13
5202428-2
;Patent No. 5202428
;    APPLICANT: SCHUBERT, DAVID
;    TITLE OF INVENTION: DNA ENCODING NEUROTROPHIC GROWTH FACTOR
;    NUMBER OF SEQUENCES: 10
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/590,359
;      FILING DATE: 27-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 541,276
;      FILING DATE: 20-JUN-1990
;SEQ ID NO:2:
;      LENGTH: 243
```

```
5202428-2

  Query Match             66.9%;  Score 874.5;  DB 6;  Length 243;
  Best Local Similarity   69.7%;  Pred. No. 9.3e-79;
  Matches  177;  Conservative   24;  Mismatches   40;  Indels   13;  Gaps    5;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              || | |   |  |||||:||||||||||:|| | ||| || |||||| |  |
Db          1 MRTPSLSLALSVLSLLVLGSGHYAAGLELNGTSSGKGEPSSGDHSAGGLVV--------- 51

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||:|  :|||||  || |:|  |||||||||||||  ||||||||||||:|| :|||| | :
Db         52 SEVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIVDDSVRVEQVIKPKENKTEGEKS 110

Qy        121 SDKPKRKKKGGKNGKNRRNR-KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGE 179
              |:||||||||||  ||  ||||  ||||||  |:|||||||||:|||:|| |||  |  |:||||
Db        111 SEKPKRKKKGGKGGKGRRNRKKKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGE 170

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVIT-VQLRRQYVRKYEGEAEER 238
              |||||:|||     ||  |||||||||  |:||| : |: :|| | ||::: |:|| |||||
Db        171 RCGEKTMKTQKKDDSDLSKIALAAIIVFVSAVSVAAIGIITAVLLRKRFFREYE-EAEER 229

Qy        239 KKLRQENGNVHAIA 252
              ::||||||  ||||
Db        230 RRLRQENGTAHAIA 243


RESULT 14
5202428-2
;Patent No. 5202428
;    APPLICANT: SCHUBERT, DAVID
;    TITLE OF INVENTION: DNA ENCODING NEUROTROPHIC GROWTH FACTOR
;    NUMBER OF SEQUENCES: 10
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/590,359
;      FILING DATE: 27-SEP-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 541,276
;      FILING DATE: 20-JUN-1990
;SEQ ID NO:2:
;      LENGTH: 243
5202428-2

  Query Match             66.9%;  Score 874.5;  DB 6;  Length 243;
  Best Local Similarity   69.7%;  Pred. No. 9.3e-79;
  Matches  177;  Conservative   24;  Mismatches   40;  Indels   13;  Gaps    5;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              || | |   |  |||||:||||||||||:|| | ||| || |||||| |  |
Db          1 MRTPSLSLALSVLSLLVLGSGHYAAGLELNGTSSGKGEPSSGDHSAGGLVV--------- 51

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||:|  :|||||  || |:|  |||||||||||||  ||||||||||||:|| :|||| | :
Db         52 SEVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIVDDSVRVEQVIKPKENKTEGEKS 110

Qy        121 SDKPKRKKKGGKNGKNRRNR-KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGE 179
              |:||||||||||  ||  ||||  ||||||  |:|||||||||:|||:|| |||  |  |:||||
Db        111 SEKPKRKKKGGKGGKGRRNRKKKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGE 170

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVIT-VQLRRQYVRKYEGEAEER 238
              |||||:|||     ||  |||||||||  |:||| : |: :|| | ||::: |:|| |||||
Db        171 RCGEKTMKTQKKDDSDLSKIALAAIIVFVSAVSVAAIGIITAVLLRKRFFREYE-EAEER 229

Qy        239 KKLRQENGNVHAIA 252
              ::||||||  ||||
Db        230 RRLRQENGTAHAIA 243


RESULT 15
US-07-885-089B-34
; Sequence 34, Application US/07885089B
; Patent No. 5830995
;  GENERAL INFORMATION:
;    APPLICANT:  Shoyab, Mohammed
;    APPLICANT:  McDonald, Vicki L.
;    APPLICANT:  Bradley, James G.
;    APPLICANT:  Plowman, Gregory D.
;    TITLE OF INVENTION:  AMPHIREGULINS:  A FAMILY OF
;    TITLE OF INVENTION:  HEPARIN-BINDING EPITHELIAL CELL GROWTH FACTORS
;    NUMBER OF SEQUENCES:  37
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Pennie & Edmonds LLP
;      STREET:  1155 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,089B
;      FILING DATE:  18-MAY-1992
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Coruzzi, Laura A.
;      REGISTRATION NUMBER:  30,742
;      REFERENCE/DOCKET NUMBER:  5624-174
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-790-9090
;      TELEFAX:  212-869-9741
;      TELEX:  66141 PENNIE
;  INFORMATION FOR SEQ ID NO:  34:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  158 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  protein
US-07-885-089B-34

  Query Match             65.4%;  Score 855;  DB 2;  Length 158;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-77;
  Matches  158;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         27 LDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEE 86
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEE 60

Qy         87 YDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPC 146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 YDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPC 120

Qy        147 NAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEK 184
              ||||||||||||||||||||||||||||||||||||||
Db        121 NAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEK 158


Search completed: February 26, 2005, 16:45:48
Job time : 44 secs
```

OM protein - protein search, using sw model

Run on: February 26, 2005, 15:47:23 ; Search time 164 Seconds
(without alignments)
594.290 Million cell updates/sec

Title: US-10-774-076-1
Perfect score: 1308
Sequence: 1 MRAPLLPPAPVVLSLLILGS.....GEAEERKKLRQENGNVHAIA 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1308	100.0	252	6	ABR47391	Abr47391 Breast ca
2	1308	100.0	252	6	ABU57614	Abu57614 Different
3	1308	100.0	252	8	ADE77063	Ade77063 Human pro
4	1308	100.0	252	8	ADH17052	Adh17052 Human amp
5	1308	100.0	252	8	ADL12853	Adl12853 Human ste
6	1308	100.0	252	8	ADJ75567	Adj75567 Marker ge
7	1308	100.0	252	8	ADP07715	Adp07715 Human sec
8	1308	100.0	252	8	ADR73587	Adr73587 Wild-type
9	1308	100.0	252	8	ADR67228	Adr67228 Human bla
10	1308	100.0	326	4	AAG75109	Aag75109 Human col
11	1295	99.0	252	1	AAP95447	Aap95447 Amphiregu
12	1226	93.7	235	7	ADC71538	Adc71538 Human NOV
13	1226	93.7	235	8	ADN33949	Adn33949 Human nov
14	1177	90.0	226	2	AAW73157	Aaw73157 Human amp
15	1004.5	76.8	247	2	AAW73160	Aaw73160 Bovine am
16	918	70.2	248	8	ADJ76293	Adj76293 Marker ge
17	914	69.9	248	2	AAW73162	Aaw73162 Mouse amp
18	874.5	66.9	243	2	AAR34705	Aar34705 JSC1 prec
19	874.5	66.9	243	2	AAW73161	Aaw73161 Rat amphi
20	855	65.4	158	2	AAW73156	Aaw73156 Soluble H
21	789	60.3	152	2	AAW73158	Aaw73158 Human amp
22	762	58.3	146	2	AAW73159	Aaw73159 Human amp
23	501	38.3	91	2	AAW37304	Aaw37304 Amphiregu
24	462	35.3	84	1	AAP90449	Aap90449 Sequence
25	459	35.1	82	2	AAW73153	Aaw73153 Human amp
26	442	33.8	82	2	AAW73154	Aaw73154 Human amp
27	440	33.6	107	1	AAP95449	Aap95449 Sequence
28	356	27.2	78	2	AAW73151	Aaw73151 Mouse amp
29	355	27.1	78	2	AAW73149	Aaw73149 Bovine am
30	347.5	26.6	79	2	AAW73150	Aaw73150 Rat amphi
31	325.5	24.9	82	1	AAP95451	Aap95451 Protein s
32	283	21.6	49	8	ADN48882	Adn48882 Amphiregu
33	259	19.8	44	6	AAE36798	Aae36798 Human amp
34	246.5	18.8	462	2	AAR91735	Aar91735 Heregulin
35	231	17.7	46	5	AAG66053	Aag66053 Mouse amp
36	216	16.5	46	3	AAB34955	Aab34955 Gene 29 h
37	211.5	16.2	208	2	AAR23998	Aar23998 EGF/HB-EH
38	211.5	16.2	208	2	AAR66190	Aar66190 Diphtheri
39	211.5	16.2	208	2	AAR80785	Aar80785 Human pre
40	211.5	16.2	208	2	AAR80787	Aar80787 Monkey pr
41	211.5	16.2	208	2	AAR92898	Aar92898 Monkey HB
42	211.5	16.2	208	2	AAR92897	Aar92897 Human HBE
43	211.5	16.2	208	3	AAY58999	Aay58999 Human hep
44	211.5	16.2	208	3	AAY90406	Aay90406 Human HBE
45	211.5	16.2	208	3	AAY90408	Aay90408 Monkey HB

ALIGNMENTS

RESULT 1
ABR47391
ID ABR47391 standard; protein; 252 AA.
XX
AC ABR47391;
XX
DT 12-JUN-2003 (first entry)
XX
DE Breast cancer associated protein sequence SEQ ID NO:13.
XX
KW Human; breast cancer; cytostatic; gene therapy.
XX
OS Homo sapiens.
XX
PN WO2003004989-A2.
XX
PD 16-JAN-2003.
XX
PF 21-JUN-2002; 2002WO-US019669.
XX
PR 21-JUN-2001; 2001US-0299887P.
PR 27-JUN-2001; 2001US-0301572P.
PR 18-JUL-2001; 2001US-0306501P.
PR 25-SEP-2001; 2001US-0325002P.
PR 05-MAR-2002; 2002US-0362585P.
PR 14-MAY-2002; 2002US-0380391P.
XX
PA (MILL-) MILLENIUM PHARM INC.
XX
PI Lillie J, Gannavarapu M, Glatt K, Hoersh S, Kamatkar S;
PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RE;
PI Bast RC, Hortobagyi GN, Pusztai L, Meric F, Sahin A, Mills GB;
XX
DR WPI; 2003-210381/20.
DR N-PSDB; ACC50082.
XX
PT Breast cancer diagnosis or treatment by comparing the level of expression
PT of a marker in a patient sample with that in the control non-breast
PT cancer sample.
XX
PS Claim 1; SEQ ID NO 13; 128pp; English.
XX
CC The present invention describes a method for assessing whether a patient
CC is afflicted with breast cancer. The method comprises comparing the level
CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and
CC ABR47386 to ABR47632) in a patient sample and the normal level of
CC expression of the marker in a control non-breast cancer sample, where a
CC significant increase in the level of expression of the marker in the

CC patient sample and the normal level is an indication that the patient is
CC afflicted with breast cancer. The breast cancer associated sequences from
CC the present invention have cytostatic activities and can be used in gene
CC therapy. The method is useful for diagnosing and treating breast cancer.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1308; DB 6; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.2e-113;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
Db 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
Db 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy 241 LRQENGNVHAIA 252
Db 241 LRQENGNVHAIA 252

RESULT 2
ABU57614
ID ABU57614 standard; protein; 252 AA.
XX
AC ABU57614;
XX
DT 09-APR-2003 (first entry)
XX
DE Differentially expressed breast cancer associated protein #1.
XX
KW Breast cancer; differential gene expression; BC-cDNA;
KW breast cancer diagnosis; breast cancer monitoring;
KW breast cancer treatment; breast cancer staging.
XX
OS Homo sapiens.
XX
PN US2002156263-A1.
XX
PD 24-OCT-2002.
XX
PF 04-OCT-2001; 2001US-00974298.
XX
PR 05-OCT-2000; 2000US-0238331P.
XX
PA (CHEN/) CHEN H.
XX
PI Chen H;
XX
DR WPI; 2003-182653/18.
XX
PT New cDNAs, which are differentially expressed in (metastatic) breast
PT cancer useful for diagnosing or staging, breast cancer, or for monitoring
PT the treatment of breast cancer in an individual.
XX
PS Example; SEQ ID NO 2; 30pp; English.
XX
CC The invention describes a combination of cDNAs (designated BC-cDNAs),
CC which are differentially expressed in breast cancer. The combination
CC includes 152 cDNA sequences, or their complements. The protein encoded by
CC any of these BC-cDNAs is useful for screening several molecules or
CC compounds to identify at least one ligand that specifically binds the
CC protein, producing or preparing polyclonal or monoclonal antibodies, or
CC purifying antibodies from a sample. The antibodies, which specifically
CC bind the protein differentially expressed in breast cancer is useful for
CC detecting the expression of a protein in a sample. The BC-cDNAs are also
CC useful for diagnosing, monitoring the treatment of, or staging, breast
CC cancer. This is the amino acid sequence of a differentially expressed
CC breast cancer associated protein. Note: The sequence data for this patent
CC did not form part of the printed specification, but was obtained in
CC electronic format directly from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=20020156263
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1308; DB 6; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.2e-113;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
Db 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
Db 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy 241 LRQENGNVHAIA 252
Db 241 LRQENGNVHAIA 252

RESULT 3
ADE77063
ID ADE77063 standard; protein; 252 AA.
XX
AC ADE77063;
XX
DT 29-JAN-2004 (first entry)
XX
DE Human protein expressed in a liver disorder #62.
XX
KW human; liver disorder; hyperlipidaemia; hypertension; type II diabetes;
KW tumour; liver; inflammatory disorder; immune response disorder;
KW high-throughput screening; differential gene expression; gene therapy.
XX
OS Homo sapiens.
XX
PN US2003108871-A1.
XX
PD 12-JUN-2003.
XX
PF 30-JUL-2001; 2001US-00919039.
XX
PR 28-JUL-2000; 2000US-0222113P.
XX
PA (KASE/) KASER M R.
XX
PI Kaser MR;
XX
DR WPI; 2004-031227/03.
DR N-PSDB; ADE77062.
XX
PT Composition comprising several cDNAs that are differentially expressed in

PT treated human C3A liver cell cultures, useful for treating liver
PT disorders.
XX
PS Claim 1; SEQ ID NO 228; 41pp; English.
XX
CC The invention relates to a composition comprising several cDNAs that are
CC differentially expressed in a liver disorder. The composition is useful
CC for treating liver disorder such as hyperlipidaemia, hypertension, type
CC II diabetes, tumours of the liver and disorders of the inflammatory and
CC immune response. The composition is useful for a high-throughput method
CC of screening several molecules or compounds to identify a ligand which
CC specifically binds a cDNA. A protein encoded by the cDNA is useful for a
CC high-throughput method for using a protein to screen several molecules or
CC compounds to identify at least one ligand which specifically binds the
CC protein which involves combining the protein encoded by the cDNA with
CC several of molecules or compounds under conditions to allow specific
CC binding, and detecting specific binding between the protein and a
CC molecule or compound, therefore identifying a ligand which specifically
CC binds the protein. The composition is useful for detecting and
CC quantifying differential gene expression, can be used in gene therapy, to
CC formulate prognosis and to design a treatment regimen and to monitor the
CC efficacy of treatment. The present sequence represents the amino acid
CC sequence of a protein encoded by a cDNA differentially expressed in a
CC liver disorder.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1308; DB 8; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.2e-113;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252
```

RESULT 4
ADH17052
ID ADH17052 standard; protein; 252 AA.
XX
AC ADH17052;
XX
DT 11-MAR-2004 (first entry)
XX
DE Human amphiregulin protein.
XX
KW tyrosine kinase activity; type 1 plasminogen activator inhibitor; PAI-1;
KW TIMP-1; tissue inhibitor of metalloproteinase 1; vinculin;
KW vascular endothelial growth factor; VEGF; placental growth factor; PLGF;
KW migration inhibitory factor; MIG; human; amphiregulin.
XX
OS Homo sapiens.
XX
PN WO2003097854-A2.
XX
PD 27-NOV-2003.
XX
PF 19-MAY-2003; 2003WO-US015711.
XX
PR 17-MAY-2002; 2002US-0380872P.
PR 24-FEB-2003; 2003US-0448874P.
PR 24-FEB-2003; 2003US-0448922P.
XX
PA (SUGE-) SUGEN INC.
XX
PI Morimoto A, Deprimo S, O'farrell A, Smolich BD, Manning WC;
PI Walter SA, Schilling JW, Cherrington J;
XX
DR WPI; 2004-042604/04.
XX
PT Determining whether a test compound inhibits tyrosine kinase activity in
PT a mammal by exposing the mammal to the test compound and measuring in the
PT mammal the level of at least one of the measured proteins or mRNA
PT transcripts.
XX
PS Disclosure; SEQ ID NO 51; 408pp; English.
XX
CC The invention relates to a novel method for determining whether a test
CC compound inhibits tyrosine kinase activity in a mammal comprising
CC measuring in the mammal the level of at least one of the proteins and/or
CC mRNA transcripts or genes for such proteins comprising type 1 plasminogen
CC activator inhibitor (PAI-1), TIMP-1 (tissue inhibitor of
CC metalloproteinase 1), vinculin, vascular endothelial growth factor
CC (VEGF), placental growth factor (PLGF), VEGF/PLGF heterodimers or
CC migration inhibitory factor (MIG), exposing the mammal to the test
CC compound and then measuring in the mammal the level of at least one of
CC the proteins and/or mRNA transcripts previously measured. The method of
CC the invention may be useful for determining whether a test compound
CC inhibits tyrosine kinase activity in a mammal. The current sequence is
CC that of the tyrosine kinase activity inhibition-related protein of the
CC invention.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1308; DB 8; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.2e-113;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252
```

RESULT 5
ADL12853
ID ADL12853 standard; protein; 252 AA.
XX
AC ADL12853;
XX
DT 06-MAY-2004 (first entry)
XX
DE Human steroid-induced C3A liver cell protein #92.
XX
KW Hepatotropic; Gene therapy; Wilson disease; liver disorder;

```
KW   steroid therapy; cirrhosis; hepatitis; human; C3A liver cell.
XX
OS   Homo sapiens.
XX
PN   US6673549-B1.
XX
PD   06-JAN-2004.
XX
PF   12-OCT-2001; 2001US-00976594.
XX
PR   12-OCT-2000; 2000US-0240409P.
XX
PA   (INCY-) INCYTE CORP.
XX
PI   Furness LM,  Buchbinder JL;
XX
DR   WPI; 2004-068610/07.
XX
PT   Combination useful for preparing a composition for treating liver
PT   disorders associated with steroid therapy, e.g., cirrhosis or hepatitis,
PT   comprises cDNAs that are differentially expressed in response to steroid
PT   treatment.
XX
PS   Disclosure; SEQ ID NO 582; 141pp; English.
XX
CC   The invention relates to a combination comprising cDNAs that are
CC   differentially expressed in response to steroid treatment. Also included
CC   are the following: a high throughput method for using a cDNA to detect
CC   differential expression of nucleic acids in a sample; and a high
CC   throughput method of screening molecules or compounds to identify a
CC   ligand that specifically binds a cDNA. The sample is from a subject with
CC   Wilson disease and comparison of a standard defines a stage of that
CC   disease. The high throughput method of screening molecules or compounds
CC   to identify a ligand that specifically binds a cDNA comprises: combining
CC   the combination with molecules or compounds under conditions to allow
CC   specific binding; and detecting specific binding between each cDNA and at
CC   least one molecule or compound. The molecules or compounds are regulatory
CC   proteins. The combination is useful for preparing a composition for
CC   treating liver disorders associated with steroid therapy, e.g., cirrhosis
CC   or hepatitis. The present sequence represents a human protein which is
CC   differentially expressed in steroid-induced C3A liver cells. Note: The
CC   sequence data for this patent did not form part of the printed
CC   specification but was obtained in electronic format directly from USPTO
CC   at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 252 AA;

  Query Match             100.0%;  Score 1308;  DB 8;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-113;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252

RESULT 6
ADJ75567
ID   ADJ75567 standard; protein; 252 AA.
XX
AC   ADJ75567;
XX
DT   20-MAY-2004  (first entry)
XX
DE   Marker gene related amino acid sequence SEQ ID NO:819.
XX
KW   bronchial asthma; chronic obstructive pulmonary disease;
KW   respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
KW   gene therapy; marker.
XX
OS   Homo sapiens.
XX
PN   EP1394274-A2.
XX
PD   03-MAR-2004.
XX
PF   04-AUG-2003; 2003EP-00254857.
XX
PR   06-AUG-2002; 2002JP-00229312.
PR   20-MAR-2003; 2003JP-00077212.
XX
PA   (GENO-) GENOX RES INC.
XX
PI   Ohtani N,  Sugita Y,  Yamaya M,  Kubo H,  Nagai H,  Izuhara K;
XX
DR   WPI; 2004-193155/19.
XX
PT   Testing for bronchial asthma or chronic obstructive pulmonary disease by
PT   comparing the expression level of a marker gene in a biological sample
PT   from a subject with the expression level of the gene in a sample from a
PT   healthy subject.
XX
PS   Example 11; SEQ ID NO 819; 241pp; English.
XX
CC   The present invention describes a method of testing for bronchial asthma
CC   or chronic obstructive pulmonary disease. The method comprises
CC   determining the expression level of a marker gene in a biological sample
CC   from a subject, comparing the expression level determined with the
CC   expression level of the marker gene in a biological sample from a healthy
CC   subject, and judging whether the subject has bronchial asthma or chronic
CC   obstructive pulmonary disease. The marker gene comprises: (a) a group of
CC   genes (S1) whose expression levels increase when respiratory epithelial
CC   cells are stimulated with interleukin-13; or (b) a group of genes (S2)
CC   whose expression levels decrease when respiratory epithelial cells are
CC   stimulated with interleukin-13. Also described: (1) a reagent (I) for
CC   testing for bronchial asthma or chronic obstructive pulmonary disease;
CC   (2) a kit for screening for a candidate compound for a therapeutic agent
CC   to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
CC   an animal model for bronchial asthma or chronic obstructive pulmonary
CC   disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
CC   method for producing an animal model for bronchial asthma or chronic
CC   obstructive pulmonary disease; (6) a therapeutic agent for bronchial
CC   asthma or chronic obstructive pulmonary disease, comprising the compound,
CC   a marker gene or an antisense nucleic acid corresponding to a portion of
CC   the marker gene, a ribozyme, a polynucleotide that suppresses the
CC   expression of the gene through an RNAi effect or an antibody recognising
CC   a protein encoded by a marker gene; and (7) a DNA chip for testing for
CC   bronchial asthma or a chronic obstructive pulmonary disease, on which a
CC   probe has been immobilised to assay a marker gene. (I) has respiratory
CC   and antiasthmatic activities, and can be used in gene therapy. The method
CC   is useful for testing for or screening for a therapeutic agent for
CC   bronchial asthma or chronic obstructive pulmonary disease. The present
CC   sequence is used in the exemplification of the present invention.
XX
SQ   Sequence 252 AA;

  Query Match             100.0%;  Score 1308;  DB 8;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-113;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252
```

```
RESULT 7
ADP07715
ID   ADP07715 standard; protein; 252 AA.
XX
AC   ADP07715;
XX
DT   12-AUG-2004  (first entry)
XX
DE   Human secreted protein, seq id 198.
XX
KW   Cytostatic; antidiabetic; anorectic; gynaecological; antipsoriatic;
KW   dermatological; antiarteriosclerotic; antiasthmatic; neuroprotective;
KW   nootropic; antiparkinsonian; nephrotropic; human; secreted protein;
KW   diagnostic; pharmaceutical; cancer; lung; oesophageal; liver; diabetes;
KW   obesity; metabolic disorder; cardiovascular disorder;
KW   reproductive disorder; psoriasis; eczema; bronchitis; cystic fibrosis;
KW   atherosclerosis; benign prostatic hyperplasia; asthma;
KW   Alzheimer's disease; Parkinson's disease; renal disorder.
XX
OS   Homo sapiens.
XX
PN   WO2004042000-A2.
XX
PD   21-MAY-2004.
XX
PF   16-MAY-2003; 2003WO-US015439.
XX
PR   17-MAY-2002; 2002US-0381592P.
PR   12-JUN-2002; 2002US-0388543P.
PR   08-AUG-2002; 2002US-0401757P.
PR   12-AUG-2002; 2002US-0402585P.
PR   13-AUG-2002; 2002US-0402799P.
PR   22-AUG-2002; 2002US-0404959P.
PR   04-OCT-2002; 2002US-0415902P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM,  Olsen H,  Baker KP,  Fiscella M,  Wei P;
PI   Birse CE,  Komatsoulis G,  Choi GH,  Moore PA,  Gupta R,  Shi Y;
XX
DR   WPI; 2004-400658/37.
DR   N-PSDB; ADP07533.
XX
PT   New human secreted polypeptides and nucleic acid molecules for
PT   diagnosing, preventing or treating disorders associated with the secreted
PT   proteins, such as cancer, diabetes, obesity, cardiovascular disorders or
PT   renal disorders.
XX
PS   Claim 1; SEQ ID NO 198; 1157pp; English.
XX
CC   The invention relates to a human secreted polypeptide for diagnosing,
CC   preventing or treating disorders associated with the secreted proteins.
CC   The polypeptides and nucleic acid molecules of the invention are useful
CC   for preparing a diagnostic or pharmaceutical composition for diagnosing
CC   or treating a medical condition. These may be used for diagnosing,
CC   preventing or treating disorders related to the human secreted proteins,
CC   such as cancer (e.g. lung, oesophageal or liver cancer), diabetes,
CC   obesity, metabolic disorders, cardiovascular disorders, reproductive
CC   disorders, psoriasis, eczema, bronchitis, cystic fibrosis,
CC   atherosclerosis, benign prostatic hyperplasia, asthma, Alzheimer's
CC   disease, Parkinson's disease or renal disorders. Sequences given in
CC   records for ADP07710-ADP07891 represent human secreted proteins of the
CC   invention.
XX
SQ   Sequence 252 AA;

  Query Match             100.0%;  Score 1308;  DB 8;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.2e-113;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252
```

```
RESULT 8
ADR73587
ID   ADR73587 standard; protein; 252 AA.
XX
AC   ADR73587;
XX
DT   04-NOV-2004  (first entry)
XX
DE   Wild-type full-length human AR polypeptide, seq id 1.
XX
KW   Antipsoriatic; cytostatic; gene therapy; amphiregulin; AR; heparin;
KW   PAR34; PAR80; HuPAR34; chimeric; humanised; antibody; cancer; psoriasis;
KW   human.
XX
OS   Homo sapiens.
XX
PN   WO2004068931-A2.
XX
PD   19-AUG-2004.
XX
PF   06-FEB-2004; 2004WO-US004176.
XX
PR   07-FEB-2003; 2003US-0445640P.
PR   30-DEC-2003; 2003US-0533901P.
XX
PA   (PROT-) PROTEIN DESIGN LABS INC.
XX
PI   Landolfi NF,  Tsurushita N,  Hinton PR,  Kumar S;
XX
DR   WPI; 2004-604310/58.
XX
PT   New antibody that competitively inhibits binding of an amphiregulin (AR)
PT   polypeptide, useful for diagnosing or treating patients having cancer and
PT   psoriasis.
```

XX
PS Example 1; SEQ ID NO 1; 99pp; English.
XX
CC The invention relates to an antibody that competitively inhibits binding
CC of an amphiregulin (AR) polypeptide to an antibody. Ar is a heparin-
CC binding glycoprotein. The antibody comprises a heavy chain variable
CC region having ammo acid sequence of at least 60% identity to a sequence
CC selected from ADR73588, ADR73590 and ADR73598, and a light chain variable
CC region having amino acid sequence of at least 60% identity to a sequence
CC selected from ADR73589, ADR73591 and ADR73600. The antibody binds to the
CC same AR epitope as that bound by an antibody selected from PAR34, PAR80,
CC or HuPAR34, and is chimeric or humanised. The antibodies are useful for
CC diagnosing or treating patients having cancer and psoriasis. The current
CC sequence represents the wild-type full-length human AR polypeptide.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1308; DB 8; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.2e-113;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
Db 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
Db 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy 241 LRQENGNVHAIA 252
Db 241 LRQENGNVHAIA 252

RESULT 9
ADR67228
ID ADR67228 standard; protein; 252 AA.
XX
AC ADR67228;
XX
DT 02-DEC-2004 (first entry)
XX
DE Human bladder cancer associated amino acid sequence.
XX
KW bladder cancer tissue; bladder cancer; cytostatic.
XX
OS Homo sapiens.
XX
PN WO2004076613-A2.
XX
PD 10-SEP-2004.
XX
PF 24-FEB-2004; 2004WO-DE000364.
XX
PR 26-FEB-2003; 2003DE-01009729.
XX
PA (HERR/) HERR A.
PA (HINZ/) HINZMANN B.
PA (DAHL/) DAHL E.
PA (STAU/) STAUB E.
PA (PILA/) PILARSKY C.
PA (SPEC/) SPECHT T.
XX
PI Herr A, Hinzmann B, Dahl E, Staub E, Pilarsky C, Specht T;
XX
DR WPI; 2004-653385/63.
XX
PT New nucleic acids, and encoded proteins, from bladder cancer tissue,
PT useful for diagnosis, treatment and in screening for specific binding
PT agents.
XX
PS Claim 2; Fig 3; 112pp; German.
XX
CC The present invention describes nucleic acids (I) associated with bladder
CC cancer tissue. Also described: (1) peptides and proteins (II) containing
CC an amino acid sequence encoded by (I); (2) a method for diagnosing
CC bladder cancer (BC), or monitoring its progression, that uses (I), (II)
CC or agents (Z) that inhibit, or bind to, (I) or (II), labelled with a
CC reporter; and (3) a method for treating BC that uses (I), (II) or (Z).
CC (I) and (II) have cytostatic activity. (I) and (II) can be used to detect
CC (and monitor progression of) bladder cancer (BC), or the risk of
CC developing it; to screen for specific binding agents (Z), and to treat
CC BC. (Z) are also useful as diagnostic and therapeutic agents. The present
CC sequence represents a human amino acid sequence associated with bladder
CC cancer, which is used in the exemplification of the present invention.
XX
SQ Sequence 252 AA;

Query Match 100.0%; Score 1308; DB 8; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.2e-113;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db 1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db 61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
Db 121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
Db 181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy 241 LRQENGNVHAIA 252
Db 241 LRQENGNVHAIA 252

RESULT 10
AAG75109
ID AAG75109 standard; protein; 326 AA.
XX
AC AAG75109;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5873.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 4.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.

```
XX
PI   Ruben SM,  Barash SC,  Birse CE,  Rosen CA;
XX
DR   WPI; 2001-235357/24.
DR   N-PSDB; AAH34514.
XX
PT   Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT   useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS   Claim 11; Page 7390-7392; 9803pp; English.
XX
CC   AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC   cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC   proteins are collectively known as colon cancer antigens. The colon
CC   cancer antigens have cytostatic activity and can be used in gene therapy
CC   and vaccine production. N and P may be used in the prevention, diagnosis
CC   and treatment of diseases associated with inappropriate P expression. For
CC   example, N and P may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of P by expressing inactive proteins or to
CC   supplement the patients own production of P. Additionally, N may be used
CC   to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC   into a host cell and culturing the cell to express the proteins. N and P
CC   can be used in the prevention, diagnosis and treatment of colorectal
CC   carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC   sequences used in the exemplification of the present invention. N.B.
CC   Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC   time of publication, meaning no sequences are present for SEQ ID NO:1027
CC   to 1052, 7921 and 7922
XX
SQ   Sequence 326 AA;

  Query Match             100.0%;  Score 1308;  DB 4;  Length 326;
  Best Local Similarity   100.0%;  Pred. No. 3.1e-113;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 134

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 194

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 254

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 314

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        315 LRQENGNVHAIA 326


RESULT 11
AAP95447
ID   AAP95447 standard; protein; 252 AA.
XX
AC   AAP95447;
XX
DT   27-AUG-2003  (revised)
DT   25-MAR-2003  (revised)
DT   10-NOV-1989  (first entry)
XX
DE   Amphiregulin precursor protein.
XX
KW   Amphiregulin; growth regulator; neoplastic disease; bone resorption;
KW   immune response; arachidonic acid cascade.
XX
OS   Human MCF 7 breast carcinoma cells.
XX
FH   Key             Location/Qualifiers
FT   Protein         101. .171
FT                   /label= amphiregulin
XX
PN   DE3902157-A.
XX
PD   27-JUL-1989.
XX
PF   25-JAN-1989;   89DE-03902157.
XX
PR   25-JAN-1988;   88US-00148327.
PR   15-APR-1988;   88US-00181884.
PR   17-JAN-1989;   89US-00297816.
XX
PA   (ONCO ) ONCOGEN.
PA   (BRIM ) BRISTOL-MYERS CO.
XX
PI   Shoyab M,  Mcdnald VL,  Plowman G,  Bradley JG;
XX
DR   WPI; 1989-221730/31.
DR   N-PSDB; AAN90373.
XX
PT   New growth regulating proteins designated amphiregulin - inhibiting
PT   tumours but stimulating some normal cells, also new precursor protein,
PT   encoding nucleic acid sequences and antibodies.
XX
PS   Disclosure; Fig 17; 78pp; German.
XX
CC   This sequence represents the precursor form of amphiregulin (ARPP). The
CC   corresp. DNA sequence is spliced to remove introns. The sequence is also
CC   processed to produce mature amphiregulin (AR). The mature protein is
CC   expressed using recombinant DNA methods and can be glycosylated or non-
CC   glycosylated. Abs can be produced which recognise an epitope of AR. AR is
CC   bifunctional growth regulator which inhibits growth of some cancers but
CC   stimulates growth of some normal cells. It is useful in the treatment
CC   diagnosis and monitoring of neoplastic diseases, for treating wounds etc.
CC   See also AAP90448, AAP90449, AAP90450, AAP90451, AAP90452, AAP90453,
CC   AAP90454, AAP90372, AAP90374, AAP90375 and AAP90376. (Updated on 25-MAR-
CC   2003 to correct PA field.) (Updated on 25-MAR-2003 to correct PI field.)
CC   (Updated on 27-AUG-2003 to correct OS field.)
XX
SQ   Sequence 252 AA;

  Query Match             99.0%;  Score 1295;  DB 1;  Length 252;
  Best Local Similarity   99.2%;  Pred. No. 3.5e-112;
  Matches  250;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEHSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              |||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPXNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252


RESULT 12
ADC71538
ID   ADC71538 standard; protein; 235 AA.
XX
```

```
AC   ADC71538;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human NOV6a protein SEQ ID NO:166.
XX
KW   human; NOVX; cytostatic; anorectic; antidiabetic; antiparkinsonian;
KW   neuroprotective; nootropic; immunomodulator; antiarteriosclerotic;
KW   nephrotropic; osteopathic; vaccine; gene therapy; cancer; obesity;
KW   diabetes; metabolic disorder; anorexia; neurodegenerative disease;
KW   Parkinson's disease; Alzheimer's disease; immune disorder;
KW   haematopoietic disorder; atherosclerosis; renal failure; hyperkalaemia;
KW   hypoglycaemia; bone disorder; wasting disorder.
XX
OS   Homo sapiens.
XX
PN   WO2003040329-A2.
XX
PD   15-MAY-2003.
XX
PF   05-NOV-2002; 2002WO-US035535.
XX
PR   05-NOV-2001; 2001US-0338626P.
PR   28-NOV-2001; 2001US-00996015.
PR   28-NOV-2001; 2001US-0333912P.
PR   29-NOV-2001; 2001US-0334300P.
PR   03-DEC-2001; 2001US-0338196P.
PR   19-APR-2002; 2002US-0373806P.
PR   16-MAY-2002; 2002US-0381043P.
PR   07-AUG-2002; 2002US-0401593P.
PR   04-NOV-2002; 2002US-00287190.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Alsobrook JP,  Anderson DW,  Boldog FL,  Burgess CE,  Chant JS;
PI   Chapoval A,  Chaudhuri A,  Edinger SR,  Eisen A,  Gangolli EA;
PI   Gerlach VL,  Guo X,  Ji W,  Khramtsov NV,  Leite MW,  Li L,  Mezes PS;
PI   Millet I,  Ooi CE,  Ort T,  Padigaru M,  Patturajan M,  Pena CEA;
PI   Rastelli L,  Rieger DK,  Quinn Senger KE,  Smithson G,  Spaderna SK;
PI   Spytek KA,  Stone DJ,  Twomlow N,  Vernet CAM,  Voss EZ,  Zerhusen BD;
PI   Zhong M;
XX
DR   WPI; 2003-441554/41.
DR   N-PSDB; ADC71537.
XX
PT   New NOVX polypeptides and nucleic acids for diagnosing, preventing or
PT   treating NOVX-associated disorders, e.g. cancer, obesity, diabetes or
PT   atherosclerosis, and for chromosome mapping, tissue typing or
PT   pharmacogenomics.
XX
PS   Claim 1; SEQ ID NO 166; 112pp; English.
XX
CC   The present invention describes novel human proteins designated NOVX,
CC   where X can be an integer of 1 to 7. The NOVX protein and nucleotide
CC   sequences have cytostatic, anorectic, antidiabetic, antiparkinsonian,
CC   neuroprotective, nootropic, immunomodulator, antiarteriosclerotic,
CC   nephrotropic and osteopathic activities, and can be used in vaccines and
CC   in gene therapy. The NOVX proteins can be used in manufacturing a
CC   medicament for treating a syndrome associated with a human disease, the
CC   disease selected from a pathology associated with the protein. The
CC   disease can be cancer. NOVX proteins, nucleic acid molecules and
CC   antibodies may also be used for diagnosing, treating or preventing other
CC   pathologies associated with aberrant NOVX expression or activity, such as
CC   obesity, diabetes, metabolic disorders, anorexia, neurodegenerative
CC   diseases (e.g. Parkinson's disease or Alzheimer's disease), immune
CC   disorders, haematopoietic disorders, atherosclerosis, renal failure,
CC   hyperkalaemia, hypoglycaemia, bone disorders and wasting disorders. The
CC   NOVX nucleic acids and proteins may also be used in chromosome mapping,
CC   tissue typing, predictive medicine and pharmacogenomics. The present
CC   sequence is used in the exemplification of the present invention.
XX
SQ   Sequence 235 AA;
```

```
  Query Match             93.7%;  Score 1226;  DB 7;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 8.8e-106;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSE 74
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSE 60

Qy         75 PSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNG 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNG 120

Qy        135 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDS 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDS 180

Qy        195 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVH 249
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVH 235
```

```
RESULT 13
ADN33949
ID   ADN33949 standard; protein; 235 AA.
XX
AC   ADN33949;
XX
DT   17-JUN-2004  (first entry)
XX
DE   Human novel protein NOV6a.
XX
KW   Human; NOVX; pancreatic polypeptide; angiopoietin; interleukin-1;
KW   endothelin-2; endozepine; amphiregulin; metallocarboxypeptidase;
KW   metabolic disorder; diabetes; obesity; infectious disease; anorexia;
KW   cancer; cancer-associated cachexia; neurodegenerative disorder;
KW   Alzheimer's disease; Parkinson's disease; immune disorder;
KW   haematopoietic disorder; dyslipidaemia; metabolic disturbance;
KW   metabolic syndrome X; wasting disorder; antibacterial agent; SNP;
KW   single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 86
FT                   /note= "May be Tyr as the result of a single nucleotide
FT                   polymorphism"
FT   Misc-difference 89
FT                   /note= "May be Gly as the result of a single nucleotide
FT                   polymorphism"
FT   Misc-difference 124
FT                   /note= "May be Gly as the result of a single nucleotide
FT                   polymorphism"
XX
PN   US2004038230-A1.
XX
PD   26-FEB-2004.
XX
PF   04-NOV-2002; 2002US-00287190.
XX
PR   05-NOV-2001; 2001US-0338626P.
PR   28-NOV-2001; 2001US-00996015.
PR   28-NOV-2001; 2001US-0333912P.
PR   29-NOV-2001; 2001US-0334300P.
PR   03-DEC-2001; 2001US-0338196P.
PR   19-APR-2002; 2002US-0373806P.
PR   16-MAY-2002; 2002US-0381043P.
PR   07-AUG-2002; 2002US-0401593P.
XX
PA   (ALSO/) ALSOBROOK J P.
PA   (ANDE/) ANDERSON D W.
PA   (BOLD/) BOLDOG F L.
PA   (BURG/) BURGESS C E.
PA   (CHAN/) CHANT J S.
```

PA (CHAP/) CHAPOVAL A.
PA (CHAU/) CHAUDHURI A.
PA (EDIN/) EDINGER S R.
PA (EISE/) EISEN A.
PA (GANG/) GANGOLLI E A.
PA (GERL/) GERLACH V.
PA (GUOX/) GUO X S.
PA (JIWW/) JI W.
PA (KHRA/) KHRAMTSOV N V.
PA (LEIT/) LEITE M W.
PA (LILL/) LI L.
PA (MEZE/) MEZES P S.
PA (MILL/) MILLET I.
PA (OOIC/) OOI C E.
PA (ORTT/) ORT T.
PA (PADI/) PADIGARU M.
PA (PATT/) PATTURAJAN M.
PA (PENA/) PENA C E A.
PA (RAST/) RASTELLI L.
PA (RIEG/) RIEGER D K.
PA (SENG/) SENGER K E Q.
PA (SMIT/) SMITHSON G.
PA (SPAD/) SPADERNA S K.
PA (SPYT/) SPYTEK K A.
PA (STON/) STONE D J.
PA (TWOM/) TWOMLOW N.
PA (VERN/) VERNET C A M.
PA (VOSS/) VOSS E Z.
PA (ZERH/) ZERHUSEN B D.
PA (ZHON/) ZHONG M.
XX
PI Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Chant JS;
PI Chapoval A, Chaudhuri A, Edinger SR, Eisen A, Gangolli EA;
PI Gerlach V, Guo XS, Ji W, Khramtsov NV, Leite MW, Li L, Mezes PS;
PI Millet I, Ooi CE, Ort T, Padigaru M, Patturajan M, Pena CEA;
PI Rastelli L, Rieger DK, Senger KEQ, Smithson G, Spaderna SK;
PI Spytek KA, Stone DJ, Twomlow N, Vernet CAM, Voss EZ, Zerhusen BD;
PI Zhong M;
XX
DR WPI; 2004-213932/20.
DR N-PSDB; ADN33948.
XX
PT Isolated NOVX polypeptides and nucleic acids, useful for preventing,
PT diagnosing and treating e.g. cancer, diabetes and Alzheimer's disease.
XX
PS Claim 2; SEQ ID NO 166; 129pp; English.
XX
CC The invention relates to 87 isolated NOVX polypeptides (NOV1-NOV5), their
CC mature forms and their encoding polynucleotides having sequence
CC similarity to pancreatic polypeptide (NOV1), angiopoietin (NOV2),
CC interleukin-1 (NOV3), endothelin-2 (NOV4), endozepine (NOV5),
CC amphiregulin (NOV6) and metallocarboxypeptidase (NOV7). Also included the
CC use of NOVX in the manufacture of a medicament for treating a human
CC disease associated with NOVX, detecting NOVX in a sample via an
CC immunoassay, identifying an agent that binds to the NOVX polypeptide,
CC modulating the activity of NOVX, a vector comprising NOVX polynucleotide,
CC a cell comprising the vector, an antibody that immunospecifically binds
CC to NOVX, detecting the polynucleotide in a sample using a hybridisation
CC assay; and producing the polypeptide comprising culturing the cell under
CC conditions that lead to expression of the polypeptide. NOVX and its
CC polynucleotide are used to prevent, diagnose or treat a medical condition
CC in human related to the aberrant expression and activity of NOVX
CC polypeptides e.g. metabolic disorders, diabetes, obesity, infectious
CC disease, anorexia, cancer, cancer-associated cachexia, neurodegenerative
CC disorders, Alzheimer's disease, Parkinson's disease, immune disorders,
CC haematopoietic disorders, and the various dyslipidaemias, metabolic
CC disturbances associated with obesity, the metabolic syndrome X and
CC wasting disorders associated with chronic diseases and various cancers.
CC They may also be used as antibacterial agents. NOVX may also be used as
CC an antigen in the production of antibodies and in assays to identify
CC modulators (agonists and antagonists) of the expression and activity of
CC NOVX. The present sequence represents a NOVX protein.
XX
SQ Sequence 235 AA;

Query Match 93.7%; Score 1226; DB 8; Length 235;
Best Local Similarity 100.0%; Pred. No. 8.8e-106;
Matches 235; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSE 74
Db 1 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSE 60

Qy 75 PSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNG 134
Db 61 PSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNG 120

Qy 135 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDS 194
Db 121 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDS 180

Qy 195 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVH 249
Db 181 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVH 235

RESULT 14
AAW73157
ID AAW73157 standard; protein; 226 AA.
XX
AC AAW73157;
XX
DT 20-JAN-1999 (first entry)
XX
DE Human amphiregulin transmembrane glycoprotein.
XX
KW Amphiregulin; AR protein; epithelial cell growth; tumour cell growth;
KW psoriasis; tumour marker; cancer diagnosis; therapy; human.
XX
OS Homo sapiens.
XX
PN US5830995-A.
XX
PD 03-NOV-1998.
XX
PF 18-MAY-1992; 92US-00885089.
XX
PR 25-JAN-1988; 88US-00148327.
PR 15-APR-1988; 88US-00181884.
PR 17-JAN-1989; 89US-00297816.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Plowman GD, Bradley JG, Shoyab M, Mcdonald VL;
XX
DR WPI; 1998-609313/51.
XX
PT Amphiregulin proteins - useful for stimulating e.g. epithelial cell
PT growth.
XX
PS Claim 11; Col 71-72; 76pp; English.
XX
CC This sequence represents the human amphiregulin (AR) transmembrane
CC glycoprotein of the invention. The invention also relates to rat, mouse
CC and bovine AR proteins. The AR proteins can be used to stimulate
CC epithelial cell growth. They may be of therapeutic use in controlling
CC tumour cell growth and in disorders such as psoriasis, and can be used as
CC a tumour marker for cancer diagnosis
XX
SQ Sequence 226 AA;

Query Match 90.0%; Score 1177; DB 2; Length 226;
Best Local Similarity 100.0%; Pred. No. 3.1e-101;
Matches 226; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 27 LDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEE 86

```
Db          1 LDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEE 60

Qy         87 YDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPC 146

Db         61 YDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPC 120

Qy        147 NAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAA 206

Db        121 NAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAA 180

Qy        207 FMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVHAIA 252

Db        181 FMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVHAIA 226
```

RESULT 15
AAW73160
ID AAW73160 standard; protein; 247 AA.
XX
AC AAW73160;
XX
DT 20-JAN-1999 (first entry)
XX
DE Bovine amphiregulin.
XX
KW Amphiregulin; AR protein; epithelial cell growth; tumour cell growth;
KW psoriasis; tumour marker; cancer diagnosis; therapy; cow.
XX
OS Bos sp.
XX
PN US5830995-A.
XX
PD 03-NOV-1998.
XX
PF 18-MAY-1992; 92US-00885089.
XX
PR 25-JAN-1988; 88US-00148327.
PR 15-APR-1988; 88US-00181884.
PR 17-JAN-1989; 89US-00297816.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Plowman GD, Bradley JG, Shoyab M, Mcdonald VL;
XX
DR WPI; 1998-609313/51.
DR N-PSDB; AAV59098.
XX
PT Amphiregulin proteins - useful for stimulating e.g. epithelial cell
PT growth.
XX
PS Example; Fig 1a; 76pp; English.
XX
CC This sequence represents the bovine amphiregulin (AR) of the invention.
CC The invention also relates to rat, mouse and human AR proteins. The AR
CC proteins can be used to stimulate epithelial cell growth. They may be of
CC therapeutic use in controlling tumour cell growth and in disorders such
CC as psoriasis, and can be used as a tumour marker for cancer diagnosis
XX
SQ Sequence 247 AA;

```
  Query Match           76.8%;  Score 1004.5;  DB 2;  Length 247;
  Best Local Similarity 77.5%;  Pred. No. 4.3e-85;
  Matches  196;  Conservative   19;  Mismatches   31;  Indels    7;  Gaps    2;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Db          1 MRAPLLPPAPVVLSLLIFGSAHYTAGLDVNDTYSGKGEPFSGDHSADRFEVTSRSEISSA 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Db         61 SETPPGGEL------SSVIDYDYAEEYDNEPQISGYIVDDSVRVEQVVKPKKNKTESEKT 114

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Db        115 SDKPKRKKKGGKNGKNRRNRKKKNLCDTEFQNFCIHGKCTFLEQLETVSCQCYPEYFGER 174

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAV-ITVQLRRQYVRKYEGEAEERK 239

Db        175 CGEKSMKTQSMVDSDLSKIALAAIAAFVSAMTFTAIAVFITILLRRRCLRGYEGVAEERK 234

Qy        240 KLRQENGNVHAIA 252

Db        235 KLRQENGNAHAVA 247
```

Search completed: February 26, 2005, 16:41:14
Job time : 167 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 26, 2005, 16:41:22 ; Search time 134 Seconds
(without alignments)
616.920 Million cell updates/sec

Title: US-10-774-076-1
Perfect score: 1308
Sequence: 1 MRAPLLPPAPVVLSLLILGS.....GEAEERKKLRQENGNVHAIA 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1308	100.0	252	9	US-09-974-298-2	Sequence 2, Appli
2	1308	100.0	252	10	US-09-919-039-228	Sequence 228, App
3	1308	100.0	252	14	US-10-177-293-13	Sequence 13, Appl
4	1308	100.0	252	14	US-10-301-822-8	Sequence 8, Appli
5	1308	100.0	252	15	US-10-440-464-51	Sequence 51, Appl
6	1308	100.0	326	14	US-10-106-698-5883	Sequence 5883, Ap
7	1226	93.7	235	15	US-10-287-190-166	Sequence 166, App
8	454	34.7	87	14	US-10-022-609-15	Sequence 15, Appl
9	320	24.5	87	14	US-10-022-609-16	Sequence 16, Appl
10	300	22.9	52	13	US-10-201-945-11	Sequence 11, Appl
11	283	21.6	49	15	US-10-609-370-14	Sequence 14, Appl
12	268	20.5	46	13	US-10-096-241-18	Sequence 18, Appl
13	263	20.1	45	9	US-09-817-647-10	Sequence 10, Appl
14	263	20.1	45	9	US-09-877-665-10	Sequence 10, Appl
15	263	20.1	45	13	US-10-136-573A-10	Sequence 10, Appl
16	263	20.1	45	14	US-10-215-862-10	Sequence 10, Appl
17	231	17.7	46	16	US-10-240-411-13	Sequence 13, Appl
18	211.5	16.2	208	14	US-10-096-327-2	Sequence 2, Appli
19	211.5	16.2	208	14	US-10-138-158-18	Sequence 18, Appl
20	211.5	16.2	208	14	US-10-189-360-5	Sequence 5, Appli
21	211.5	16.2	208	14	US-10-189-360-7	Sequence 7, Appli
22	211.5	16.2	208	14	US-10-395-540-2	Sequence 2, Appli
23	211.5	16.2	208	15	US-10-428-487-38	Sequence 38, Appl
24	211.5	16.2	208	17	US-10-868-577A-22	Sequence 22, Appl
25	211.5	16.2	208	17	US-10-868-577A-46	Sequence 46, Appl
26	211.5	16.2	208	17	US-10-868-577A-51	Sequence 51, Appl
27	197	15.1	46	8	US-08-736-019-153	Sequence 153, App
28	184.5	14.1	208	14	US-10-189-360-8	Sequence 8, Appli
29	166	12.7	77	14	US-10-189-360-6	Sequence 6, Appli
30	166	12.7	89	14	US-10-106-698-5940	Sequence 5940, Ap
31	163	12.5	86	14	US-10-022-609-17	Sequence 17, Appl
32	145	11.1	336	9	US-09-780-669-940	Sequence 940, App
33	145	11.1	336	9	US-09-822-827-940	Sequence 940, App
34	145	11.1	336	9	US-09-895-793-940	Sequence 940, App
35	145	11.1	336	9	US-09-895-814-940	Sequence 940, App
36	145	11.1	336	13	US-10-012-896-940	Sequence 940, App
37	145	11.1	336	14	US-10-144-678A-940	Sequence 940, App
38	145	11.1	336	14	US-10-294-025-940	Sequence 940, App
39	145	11.1	347	9	US-09-759-143-590	Sequence 590, App
40	145	11.1	347	9	US-09-780-669-590	Sequence 590, App
41	145	11.1	347	9	US-09-822-827-590	Sequence 590, App
42	145	11.1	347	9	US-09-895-793-590	Sequence 590, App
43	145	11.1	347	9	US-09-895-814-590	Sequence 590, App
44	145	11.1	347	13	US-10-012-896-590	Sequence 590, App
45	145	11.1	347	14	US-10-144-678A-590	Sequence 590, App

ALIGNMENTS

RESULT 1
US-09-974-298-2
; Sequence 2, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; CURRENT FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 2
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 3478236CD1
US-09-974-298-2

Query Match 100.0%; Score 1308; DB 9; Length 252;
Best Local Similarity 100.0%; Pred. No. 2.9e-110;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
```

```
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252

Db        241 LRQENGNVHAIA 252


RESULT 2
US-09-919-039-228
; Sequence 228, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
;  APPLICANT: Kaser, Matthew R.
;  TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
;  FILE REFERENCE: PA-0035 US
;  CURRENT APPLICATION NUMBER: US/09/919,039
;  CURRENT FILING DATE:  2002-09-09
;  PRIOR APPLICATION NUMBER: 60/222,113
;  PRIOR FILING DATE: 2000-07-28
;  NUMBER OF SEQ ID NOS: 401
;  SOFTWARE: PERL Program
; SEQ ID NO 228
;   LENGTH: 252
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030108871A1 3478236CD1
US-09-919-039-228

  Query Match             100.0%;  Score 1308;  DB 10;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-110;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252

Db        241 LRQENGNVHAIA 252


RESULT 3
US-10-177-293-13
; Sequence 13, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT:  Glatt, Karen
;  APPLICANT:  Zhao, Xumei
;  APPLICANT:  Gannavarpu, Manjula
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Mertens, Maureen
;  APPLICANT:  Myer, Vic
;  APPLICANT:  Wang, Youzhen
;  APPLICANT:  Xu, Yongyao
;  APPLICANT:  Hoersch, Sebastian
;  APPLICANT:  Monahan, John
;  APPLICANT:  Meyers, Rachel E.
;  APPLICANT:  Bast Jr., Robert C.
;  APPLICANT:  Hortobagyi, Gabriel N.
;  APPLICANT:  Pusztai, Lajos
;  APPLICANT:  Meric, Funda
;  APPLICANT:  Sahin, Aysegul
;  APPLICANT:  Mills, Gordon B.
;  TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
;  TITLE OF INVENTION:  PREVENTION, AND THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-038
;  CURRENT APPLICATION NUMBER: US/10/177,293
;  CURRENT FILING DATE:  2002-06-21
;  PRIOR APPLICATION NUMBER: US 60/299,887
;  PRIOR FILING DATE: 2001-06-21
;  PRIOR APPLICATION NUMBER: US 60/301,572
;  PRIOR FILING DATE: 2001-06-27
;  PRIOR APPLICATION NUMBER: US 60/306,501
;  PRIOR FILING DATE: 2001-07-18
;  PRIOR APPLICATION NUMBER: US 60/325,002
;  PRIOR FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 60/362,585
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/xxx,xxx
;  PRIOR FILING DATE: 2002-05-14
;  NUMBER OF SEQ ID NOS: 506
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 13
;   LENGTH: 252
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-177-293-13

  Query Match             100.0%;  Score 1308;  DB 14;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-110;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252

Db        241 LRQENGNVHAIA 252


RESULT 4
US-10-301-822-8
; Sequence 8, Application US/10301822
; Publication No. US20030148410A1
; GENERAL INFORMATION:
;  APPLICANT: Millennium Pharmaceuticals, Inc.
;  APPLICANT:  Berger, Allison
;  APPLICANT:  Guillemette, Tracy L.
;  APPLICANT:  Kamatkar, Shubhangi
;  APPLICANT:  Schlegel, Robert
;  APPLICANT:  Monahan, John E.
;  APPLICANT:  Thibodeau, Stephen N.
;  APPLICANT:  Burgart, Lawrence J.
```

```
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
;  TITLE OF INVENTION:  METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION:  THERAPY OF COLON CANCER
;  FILE REFERENCE: MPM01-029P2RNM
;  CURRENT APPLICATION NUMBER: US/10/301,822
;  CURRENT FILING DATE:  2002-11-21
;  PRIOR APPLICATION NUMBER: US 60/339,971
;  PRIOR FILING DATE: 2001-12-10
;  PRIOR APPLICATION NUMBER: US 60/361,978
;  PRIOR FILING DATE: 2002-03-05
;  PRIOR APPLICATION NUMBER: US 60/381,988
;  PRIOR FILING DATE: 2002-05-20
;  NUMBER OF SEQ ID NOS: 228
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
;   LENGTH: 252
;   TYPE: PRT
;   ORGANISM: Homo Sapiens
US-10-301-822-8

  Query Match             100.0%;  Score 1308;  DB 14;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-110;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252


RESULT 5
US-10-440-464-51
; Sequence 51, Application US/10440464
; Publication No. US20040018528A1
; GENERAL INFORMATION:
;  APPLICANT: DEPRIMO, SAMUEL
;  APPLICANT:  O'FARRELL, ANNE-MARIE
;  APPLICANT:  MORIMOTO, ALYSSA
;  APPLICANT:  SMOLICH, BEVERLY
;  APPLICANT:  MANNING, WILLIAM
;  APPLICANT:  WALTER, SARAH
;  APPLICANT:  CHERRINGTON, JULIE
;  APPLICANT:  SCHILLING, JIM
;  TITLE OF INVENTION: NOVEL BIOMARKERS OF TYROSINE KINASE INHIBITOR EXPOSURE
;  TITLE OF INVENTION:  AND ACTIVITY IN MAMMALS
;  FILE REFERENCE: 038602/1592
;  CURRENT APPLICATION NUMBER: US/10/440,464
;  CURRENT FILING DATE:  2003-05-19
;  PRIOR APPLICATION NUMBER: 60/380,872
;  PRIOR FILING DATE: 2002-05-17
;  PRIOR APPLICATION NUMBER: 60/448,922
;  PRIOR FILING DATE: 2003-02-24
;  PRIOR APPLICATION NUMBER: 60/448,874
;  PRIOR FILING DATE: 2003-02-24
;  NUMBER OF SEQ ID NOS: 185
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 51
;   LENGTH: 252
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-440-464-51

  Query Match             100.0%;  Score 1308;  DB 15;  Length 252;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-110;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252


RESULT 6
US-10-106-698-5883
; Sequence 5883, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
;  APPLICANT: Ruben et al.
;  TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptid
;  FILE REFERENCE: PA005P1
;  CURRENT APPLICATION NUMBER: US/10/106,698
;  CURRENT FILING DATE:  2002-03-27
;  PRIOR APPLICATION NUMBER: PCT/US00/26524
;  PRIOR FILING DATE: 2000-09-28
;  PRIOR APPLICATION NUMBER: US 60/157,137
;  PRIOR FILING DATE: 1999-09-29
;  PRIOR APPLICATION NUMBER: US 60/163,280
;  PRIOR FILING DATE: 1999-11-03
;  NUMBER OF SEQ ID NOS: 8564
;  SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5883
;   LENGTH: 326
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-106-698-5883

  Query Match             100.0%;  Score 1308;  DB 14;  Length 326;
  Best Local Similarity   100.0%;  Pred. No. 4.1e-110;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         75 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 134

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        135 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 194

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        195 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 254

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        255 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 314
```

```
Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        315 LRQENGNVHAIA 326


RESULT 7
US-10-287-190-166
; Sequence 166, Application US/10287190
; Publication No. US20040038230A1
; GENERAL INFORMATION:
;  APPLICANT: Alsobrook II, John P. et al.
;  TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
;  FILE REFERENCE: 21402-780B
;  CURRENT APPLICATION NUMBER: US/10/287,190
;  CURRENT FILING DATE:  2003-04-02
;  PRIOR APPLICATION NUMBER: 09/996,015
;  PRIOR FILING DATE: 2001-11-28
;  PRIOR APPLICATION NUMBER: 60/338626
;  PRIOR FILING DATE: 2001-11-05
;  PRIOR APPLICATION NUMBER: 60/373806
;  PRIOR FILING DATE: 2002-04-19
;  PRIOR APPLICATION NUMBER: 60/338196
;  PRIOR FILING DATE: 2001-12-03
;  PRIOR APPLICATION NUMBER: 60/333912
;  PRIOR FILING DATE: 2001-11-28
;  PRIOR APPLICATION NUMBER: 60/381043
;  PRIOR FILING DATE: 2002-05-16
;  PRIOR APPLICATION NUMBER: 60/401593
;  PRIOR FILING DATE: 2002-08-07
;  PRIOR APPLICATION NUMBER: 60/334300
;  PRIOR FILING DATE: 2001-11-29
;  NUMBER OF SEQ ID NOS: 194
;  SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 166
;   LENGTH: 235
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-287-190-166

  Query Match             93.7%;  Score 1226;  DB 15;  Length 235;
  Best Local Similarity   100.0%;  Pred. No. 7.3e-103;
  Matches  235;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         15 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSE 74
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEISPVSEMPSSSE 60

Qy         75 PSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNG 134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 PSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNG 120

Qy        135 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDS 194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDS 180

Qy        195 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVH 249
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNVH 235


RESULT 8
US-10-022-609-15
; Sequence 15, Application US/10022609
; Publication No. US20030023035A1
;   GENERAL INFORMATION:
;        APPLICANT: Vandlen, Richard L.
;                   Holmes, William E.
;        TITLE OF INVENTION: Structure, Production and Use of
;                            Heregulin 2 Ligands
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 460 Point San Bruno Blvd
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: patin (Genentech)
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/10/022,609
;             FILING DATE: 17-Dec-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: US/08/440,401
;             FILING DATE: 12-MAY-1995
;             APPLICATION NUMBER: 08/330161
;             FILING DATE: 25-OCT-1994
;             APPLICATION NUMBER: 08/035430
;             FILING DATE: 22-MAR-1993
;             APPLICATION NUMBER: 07/705256
;             FILING DATE: 24-MAY-1991
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Lee, Wendy M.
;             REGISTRATION NUMBER: 00,000
;             REFERENCE/DOCKET NUMBER: 712C3
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 415/225-1994
;             TELEFAX: 415/952-9881
;             TELEX: 910/371-7168
;   INFORMATION FOR SEQ ID NO: 15:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 87 amino acids
;             TYPE: amino acid
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-022-609-15

  Query Match             34.7%;  Score 454;  DB 14;  Length 87;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-33;
  Matches   87;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        141 KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIA 200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIA 60

Qy        201 LAAIAAFMSAVILTAVAVITVQLRRQY 227
              |||||||||||||||||||||||||||
Db         61 LAAIAAFMSAVILTAVAVITVQLRRQY 87


RESULT 9
US-10-022-609-16
; Sequence 16, Application US/10022609
; Publication No. US20030023035A1
;   GENERAL INFORMATION:
;        APPLICANT: Vandlen, Richard L.
;                   Holmes, William E.
;        TITLE OF INVENTION: Structure, Production and Use of
;                            Heregulin 2 Ligands
;        NUMBER OF SEQUENCES: 17
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Genentech, Inc.
;             STREET: 460 Point San Bruno Blvd
;             CITY: South San Francisco
;             STATE: California
;             COUNTRY: USA
;             ZIP: 94080
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: patin (Genentech)
```

```
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/022,609
;            FILING DATE: 17-Dec-2001
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/08/440,401
;            FILING DATE: 12-MAY-1995
;            APPLICATION NUMBER: 08/330161
;            FILING DATE: 25-OCT-1994
;            APPLICATION NUMBER: 08/035430
;            FILING DATE: 22-MAR-1993
;            APPLICATION NUMBER: 07/705256
;            FILING DATE: 24-MAY-1991
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Lee, Wendy M.
;            REGISTRATION NUMBER: 00,000
;            REFERENCE/DOCKET NUMBER: 712C3
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 415/225-1994
;            TELEFAX: 415/952-9881
;            TELEX: 910/371-7168
;   INFORMATION FOR SEQ ID NO: 16:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 87 amino acids
;            TYPE: amino acid
;            TOPOLOGY: linear
;       SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-022-609-16

  Query Match             24.5%;  Score 320;  DB 14;  Length 87;
  Best Local Similarity   68.6%;  Pred. No. 2.7e-21;
  Matches   59;  Conservative   10;  Mismatches   17;  Indels    0;  Gaps    0;

Qy        141 KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIA 200
              |||||| |:|||||||||||:|||:|| ||| | |:||||||||||:|||    || |||||
Db          1 KKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGERCGEKTMKTQKKDDSDLSKIA 60

Qy        201 LAAIAAFMSAVILTAVAVITVQLRRQ 226
              |||| |:||| : |: :|| | |:
Db         61 LAAIIVFVSAVSVAAIGIITAVLLRK 86


RESULT 10
US-10-201-945-11
; Sequence 11, Application US/10201945
; Publication No. US20020188110A1
;   GENERAL INFORMATION:
;       APPLICANT: Meissner, Paul S.
;                  Fuldner, Rebecca A.
;                  Adams, Mark D.
;       TITLE OF INVENTION: Transforming Growth Factor Alpha HI
;       NUMBER OF SEQUENCES: 15
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: MD
;            COUNTRY: USA
;            ZIP: 20850
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/10/201,945
;            FILING DATE: 25-Jul-2002
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: US/09/471,377
;            FILING DATE: 23-Dec-1999
;            APPLICATION NUMBER: 08/915,096
;            FILING DATE: <Unknown>
;            APPLICATION NUMBER: US 08/208,008
;            FILING DATE: 08-MAR-1994
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Brookes, A. Anders
;            REGISTRATION NUMBER: 36,373
;            REFERENCE/DOCKET NUMBER: PF110D1
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 301-309-8504
;            TELEX: 301-309-8439
;   INFORMATION FOR SEQ ID NO: 11:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 52 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: protein
;       SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-201-945-11

  Query Match             22.9%;  Score 300;  DB 13;  Length 52;
  Best Local Similarity   100.0%;  Pred. No. 9.1e-20;
  Matches   52;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        135 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 KNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 52


RESULT 11
US-10-609-370-14
; Sequence 14, Application US/10609370
; Publication No. US20040048295A1
; GENERAL INFORMATION:
;  APPLICANT: Young et al.
;  TITLE OF INVENTION: Heregulin-Like Factor
;  FILE REFERENCE: PF383D1
;  CURRENT APPLICATION NUMBER: US/10/609,370
;  CURRENT FILING DATE:  2003-07-01
;  PRIOR APPLICATION NUMBER: 09/097,681
;  PRIOR FILING DATE: 1998-06-16
;  PRIOR APPLICATION NUMBER: 60/049,942
;  PRIOR FILING DATE: 1997-06-17
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
;   LENGTH: 49
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-609-370-14

  Query Match             21.6%;  Score 283;  DB 15;  Length 49;
  Best Local Similarity   100.0%;  Pred. No. 2.9e-18;
  Matches   49;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        140 RKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKT 188
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 RKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKT 49


RESULT 12
US-10-096-241-18
; Sequence 18, Application US/10096241
; Publication No. US20020127594A1
;   GENERAL INFORMATION:
;       APPLICANT: Gearing, David P.
;                  Busfield, Samantha J.
;       TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;                           AND USES THEREFOR
;       NUMBER OF SEQUENCES: 33
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Fish & Richardson P.C.
;            STREET: 225 Franklin Street
;            CITY: Boston
```

; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/096,241
; FILING DATE: 12-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 18:
US-10-096-241-18

Query Match 20.5%; Score 268; DB 13; Length 46;
Best Local Similarity 100.0%; Pred. No. 6.2e-17;
Matches 46; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 141 KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 186
||||||||||||||||||||||||||||||||||||||||||||||
Db 1 KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 46

RESULT 13
US-09-817-647-10
; Sequence 10, Application US/09817647
; Patent No. US20020082229A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/817,647
; FILING DATE: 26-Mar-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/107,979
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hAR.egf
; LOCATION: 1-45
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-817-647-10

Query Match 20.1%; Score 263; DB 9; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.7e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 142 KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 186
|||||||||||||||||||||||||||||||||||||||||||||
Db 1 KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 45

RESULT 14
US-09-877-665-10
; Sequence 10, Application US/09877665
; Patent No. US20020164680A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related
; Ligands and Uses Therefor
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 1 DNA Way
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/877,665
; FILING DATE: 08-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/109,206
; FILING DATE: 30-Jun-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Conley, Deirdre L.
; REGISTRATION NUMBER: 36,487
; REFERENCE/DOCKET NUMBER: P1084R1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650/225-2066
; TELEFAX: 650/952-9881
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 45 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; FEATURE:
; NAME/KEY: hAR.egf
; LOCATION: 1-45
; IDENTIFICATION METHOD:
; OTHER INFORMATION:
; SEQUENCE DESCRIPTION: SEQ ID NO: 10:
US-09-877-665-10

Query Match 20.1%; Score 263; DB 9; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.7e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        142 KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 186
              |||||||||||||||||||||||||||||||||||||||||||||
Db          1 KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 45
```

RESULT 15
US-10-136-573A-10
; Sequence 10, Application US/10136573A
; Publication No. US20020161200A1
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J.
; APPLICANT: Mark, Melanie Rose
; APPLICANT: Zhang, Dong Xiao
; TITLE OF INVENTION: ErbB Receptor-Specific Neuregulin Related Ligands and
; TITLE OF INVENTION: Uses Therefor
; FILE REFERENCE: P1084R1C2
; CURRENT APPLICATION NUMBER: US/10/136,573A
; CURRENT FILING DATE: 2002-04-29
; PRIOR APPLICATION NUMBER: US 09/480,977
; PRIOR FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: US 08/899,437
; PRIOR FILING DATE: 1997-07-24
; PRIOR APPLICATION NUMBER: US 60/052,019
; PRIOR FILING DATE: 1997-07-09
; NUMBER OF SEQ ID NOS: 23
; SEQ ID NO 10
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-136-573A-10

Query Match 20.1%; Score 263; DB 13; Length 45;
Best Local Similarity 100.0%; Pred. No. 1.7e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        142 KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 186
              |||||||||||||||||||||||||||||||||||||||||||||
Db          1 KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 45
```

Search completed: February 26, 2005, 16:56:18
Job time : 136 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 26, 2005, 16:28:46 ; Search time 40 Seconds
(without alignments)
606.166 Million cell updates/sec

Title: US-10-774-076-1
Perfect score: 1308
Sequence: 1 MRAPLLPPAPVVLSLLILGS.....GEAEERKKLRQENGNVHAIA 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1308	100.0	252	1	A34702	amphiregulin precu
2	918	70.2	248	1	JH0612	amphiregulin precu
3	874.5	66.9	243	2	S13296	Schwannoma-derived
4	211.5	16.2	208	1	A38432	heparin-binding EG
5	211.5	16.2	208	1	A41914	diptheria toxin re
6	185.5	14.2	208	1	JC1410	heparin-binding EG
7	184.5	14.1	208	1	JC1409	heparin-binding EG
8	162	12.4	84	2	S27162	heparin-binding ep
9	144.5	11.0	380	2	G01639	transmembrane prot
10	143	10.9	346	2	T46914	hypothetical prote
11	137	10.5	141	2	S58117	EGF-like growth fa
12	119.5	9.1	177	2	A37408	betacellulin precu
13	117	8.9	860	2	JC5702	ErbB kinase activa
14	115	8.8	272	2	B71618	merozoite surface
15	114.5	8.8	1207	1	EGHU	epidermal growth f
16	114	8.7	850	2	JC5700	ErbB kinase activa
17	113.5	8.7	162	2	S68401	epiregulin precurs
18	112	8.6	1001	2	S30385	G9a protein - huma
19	109	8.3	159	1	S27195	transforming growt
20	108	8.3	82	2	A61040	transforming growt
21	108	8.3	160	2	S39795	transforming growt
22	107	8.2	121	2	I58134	transforming growt
23	107	8.2	160	1	WFHU1	transforming growt
24	106	8.1	159	1	WFRT1	transforming growt
25	106	8.1	159	2	I57497	transforming growt
26	106	8.1	868	2	JC5701	ErbB kinase activa
27	104.5	8.0	178	2	JC1467	betacellulin precu
28	104	8.0	436	2	D88826	protein lin-3 [imp
29	104	8.0	438	2	S28263	lin-3 protein prec
30	104	8.0	637	2	C43273	heregulin precurso
31	104	8.0	1670	2	T06754	DNA-directed RNA p
32	103.5	7.9	636	2	I61718	neu differentiatio
33	103.5	7.9	1133	1	EGRT	epidermal growth f
34	102.5	7.8	511	2	JC1404	CDEI-box DNA-bindi
35	102	7.8	661	2	T42754	hypothetical prote
36	101	7.7	639	2	I61719	neu differentiatio
37	100.5	7.7	296	2	A56943	sensory/motor neur
38	100.5	7.7	662	2	I61722	neu differentiatio
39	100	7.6	582	2	S69720	hypothetical prote
40	99	7.6	442	2	H86276	F14L17.4 protein -
41	99	7.6	645	2	B43273	heregulin, splice
42	98	7.5	140	2	T30766	growth factor - va
43	97.5	7.5	462	2	I38404	neu differentiatio
44	97.5	7.5	640	2	A43273	heregulin precurso
45	96.5	7.4	1217	1	EGMSMG	epidermal growth f

ALIGNMENTS

RESULT 1
A34702
amphiregulin precursor [validated] - human
C;Species: Homo sapiens (man)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: A34702; B34702; A30057; A31230; A39788; A30056
R;Plowman, G.D.; Green, J.M.; McDonald, V.L.; Neubauer, M.G.; Disteche, C.M., To
Mol. Cell. Biol. 10, 1969-1981, 1990
A;Title: The amphiregulin gene encodes a novel epidermal growth factor-related p
A;Reference number: A34702; MUID:90220581; PMID:2325643
A;Accession: A34702
A;Molecule type: DNA
A;Residues: 1-252 <PL1>
A;Cross-references: UNIPROT:P15514; GB:M30698
A;Accession: B34702
A;Molecule type: mRNA
A;Residues: 1-252 <PL2>
A;Cross-references: GB:M30704; NID:g179039; PIDN:AAA51781.1; PID:g179040
R;Shoyab, M.; Plowman, G.D.; McDonald, V.L.; Bradley, J.G.; Todaro, G.J.
Science 243, 1074-1076, 1989
A;Title: Structure and function of human amphiregulin: a member of the epiderma
A;Reference number: A30057; MUID:89162036; PMID:2466334
A;Accession: A30057
A;Molecule type: protein
A;Residues: 101-112,'D',114-184 <SH1>
R;Shoyab, M.; McDonald, V.L.; Bradley, J.G.; Todaro, G.J.
Proc. Natl. Acad. Sci. U.S.A. 85, 6528-6532, 1988
A;Title: Amphiregulin: a bifunctional growth-modulating glycoprotein produced
A;Reference number: A94205; MUID:88320474; PMID:3413110
A;Accession: A31230
A;Molecule type: protein
A;Residues: 101-112,'D',114-131 <SH2>
R;Cook, P.W.; Mattox, P.A.; Keeble, W.W.; Pittelkow, M.R.; Plowman, G.D.; Sho
Mol. Cell. Biol. 11, 2547-2557, 1991
A;Title: A heparin sulfate-regulated human keratinocyte autocrine factor is s
A;Reference number: A39788; MUID:91203875; PMID:2017164
A;Accession: A39788
A;Molecule type: protein
A;Residues: 101-126,<COO>
C;Comment: Amphiregulin is a secreted protein that is released from the tran
C;Genetics:
A;Gene: GDB:AREG
A;Cross-references: GDB:119697; OMIM:104640
A;Map position: 4q13-4q21
A;Introns: 21/1; 104/1; 171/1; 222/2
C;Superfamily: amphiregulin; EGF homology
C;Keywords: extracellular protein; glycoprotein; growth factor; transmembra
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-100/Domain: propeptide #status predicted <PRO>
F;101-184/Product: amphiregulin long form #status experimental <MAL>
F;107-184/Product: amphiregulin short form #status experimental <MAS>
F;146-181/Domain: EGF homology <EGF>

F;199-221/Domain: transmembrane #status predicted <TM1>
F;222-252/Domain: intracellular #status predicted <INT>
F;30,113,119/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;146-159,154-170,172-181/Disulfide bonds: #status predicted

Query Match 100.0%; Score 1308; DB 1; Length 252;
Best Local Similarity 100.0%; Pred. No. 8e-95;
Matches 252; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
Db        241 LRQENGNVHAIA 252
```

RESULT 2
JH0612
amphiregulin precursor - mouse
N;Alternate names: Schwannoma-derived growth factor precursor
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004
C;Accession: JH0612; A56830
R;Sonoda, H.; Yamaguchi, T.; Watanabe, S.
Biochem. Biophys. Res. Commun. 185, 103-109, 1992
A;Title: Androgen-responsive expression and mitogenic activity of schwannoma-derived grc
A;Reference number: JH0612; MUID:92287078; PMID:1318038
A;Accession: JH0612
A;Molecule type: mRNA
A;Residues: 1-248 <SON>
A;Cross-references: UNIPROT:P31955; GB:D12648; DDBJ:D01182; NID:g220599; PIDN:BAA02169.1
A;Experimental source: SC2G cell
R;Das, S.K.; Chakraborty, I.; Paria, B.C.; Wang, X.N.; Plowman, G.; Dey, S.K.
Mol. Endocrinol. 9, 691-705, 1995
A;Title: Amphiregulin is an implantation-specific and progesterone-regulated gene in the
A;Reference number: A56830; MUID:96174905; PMID:8592515
A;Accession: A56830
A;Molecule type: mRNA
A;Residues: 1-248 <DAS>
A;Cross-references: GB:L41352; NID:g845598; PIDN:AAB00472.1; PID:g845599
A;Note: authors translated the codon CCA for residue 84 as Glu
C;Comment: Amphiregulin is a secreted protein that is released from the transmembrane pr
C;Superfamily: amphiregulin; EGF homology
C;Keywords: extracellular protein; glycoprotein; growth factor; transmembrane protein
F;1-19/Domain: signal sequence #status predicted <SIG>
F;20-93/Domain: propeptide #status predicted <PRO>
F;94-177/Product: amphiregulin long form #status predicted <MAL>
F;100-177/Product: amphiregulin short form #status predicted <MAS>
F;139-174/Domain: EGF homology <EGF>
F;192-217/Domain: transmembrane #status predicted <TM1>
F,218-248/Domain: intracellular #status predicted <INT>
F,106/Binding site: carbohydrate (Asn) (covalent) #status predicted
F,139-152,147-163,165-174/Disulfide bonds: #status predicted

Query Match 70.2%; Score 918; DB 1; Length 248;
Best Local Similarity 69.4%; Pred. No. 2.1e-64;
Matches 177; Conservative 27; Mismatches 41; Indels 10; Gaps 3;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db          1 MRTPLLPLARSVLLLLVLGSGHYAAALELNDPSSGKGESLSGDHSAGGL------ELSVG 54

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db         55 REVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIIDDSVRVEQVIKPKKNKTEGEKS 113

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
Db        114 TEKPKRKKKGGKNGKGRRNKKKKNPCTAKFQNFCIHGECRYIENLEVVTCNCHQDYFGER 173

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVA---VITVQLRRQYVRKYEGEAEE 237
Db        174 CGEKSMKTHSEDDKDLSKIAVVAVTIFVSAIILAAIGIGIVITVHLWKRYFREYEGETEE 233

Qy        238 RKKLRQENGNVHAIA 252
Db        234 RRRLRQENGTVHAIA 248
```

RESULT 3
S13296
Schwannoma-derived growth factor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S13296
R;Kimura, H.; Fischer, W.H.; Schubert, D.
Nature 348, 257-260, 1990
A;Title: Structure, expression and function of a schwannoma-derived growth facto
A;Reference number: S13296; MUID:91043095; PMID:2234093
A;Accession: S13296
A;Molecule type: mRNA
A;Residues: 1-243 <KIM>
A;Cross-references: UNIPROT:P24338; GB:X55183; NID:g49595; PIDN:CAA38967.1; PID:g
C;Superfamily: amphiregulin; EGF homology
C;Keywords: transmembrane protein
F;137-172/Domain: EGF homology <EGF>

Query Match 66.9%; Score 874.5; DB 2; Length 243;
Best Local Similarity 69.7%; Pred. No. 5.1e-61;
Matches 177; Conservative 24; Mismatches 40; Indels 13; Gaps 5;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
Db          1 MRTPSLSLALSVLSLLVLGSGHYAAGLELNGTSSGKGEPSSGDHSAGGLVV--------- 51

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
Db         52 SEVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIVDDSVRVEQVIKPKENKTEGEKS 110

Qy        121 SDKPKRKKKGGKNGKNRRNR-KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGE 179
Db        111 SEKPKRKKKGGKGGKGRRNRKKKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGE 170

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVIT-VQLRRQYVRKYEGEAEER 238
Db        171 RCGEKTMKTQKKDDSDLSKIALAAIIVFVSAVSVAAIGIITAVLLRKRFFREYE-EAEER 229

Qy        239 KKLRQENGNVHAIA 252
Db        230 RRLRQENGTAHAIA 243
```

RESULT 4
A38432
heparin-binding EGF-like growth factor precursor - human
C;Species: Homo sapiens (man)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A38432; A37300
R;Higashiyama, S.; Abraham, J.A.; Miller, J.; Fiddes, J.C.; Klagsbrun, M.
Science 251, 936-939, 1991
A;Title: A heparin-binding growth factor secreted by macrophage-like cells that is re
A;Reference number: A38432; MUID:91157008; PMID:1840698

A;Accession: A38432
A;Molecule type: mRNA
A;Residues: 1-208 <HIG>
A;Cross-references: UNIPROT:Q99075; GB:M60278; NID:g183866; PIDN:AAA35956.1; PID:g183867
R;Higashiyama, S.; Lau, K.; Besner, G.E.; Abraham, J.A.; Klagsbrun, M.
J. Biol. Chem. 267, 6205-6212, 1992
A;Title: Structure of heparin-binding EGF-like growth factor. Multiple forms, primary st
A;Reference number: A37300; MUID:92210596; PMID:1556128
A;Accession: A37300
A;Molecule type: protein
A;Residues: 63-74,'X',76-84,'X',86-148 <HI2>
C;Genetics:
A;Gene: GDB:DTR; DTS; HEGFL
A;Cross-references: GDB:119853; OMIM:126150
A;Map position: 5q23-5q23
C;Superfamily: heparin-binding EGF-like growth factor; EGF homology
C;Keywords: heparin binding; transmembrane protein
F;108-143/Domain: EGF homology <EGF>

Query Match 16.2%; Score 211.5; DB 1; Length 208;
Best Local Similarity 36.7%; Pred. No. 2.3e-09;
Matches 55; Conservative 23; Mismatches 59; Indels 13; Gaps 4;

Qy 100 DSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGEC 159
Db 70 DLLRVTLSSKPQALAT--PNKEEHGKRKKKGKGLG------KKRDPCLRKYKDFCIHGEC 121

Qy 160 KYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVI 219
Db 122 KYVKELRAPSCICHPGYHGERCHGLSLPVENRLYTYDHTTILAVVAVVLSSVCL---LVI 178

Qy 220 TVQLRRQYVRK--YEGEAEERKKLRQENGN 247
Db 179 VGLLMFRYHRRGGYDVENEEKVKLGMTNSH 208

RESULT 5
A41914
diptheria toxin receptor precursor - green monkey
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A41914; JC4687
R;Naglich, J.G.; Metherall, J.E.; Russell, D.W.; Eidels, L.
Cell 69, 1051-1061, 1992
A;Title: Expression cloning of a diphtheria toxin receptor: identity with a heparin-bind
A;Reference number: A41914; MUID:92298386; PMID:1606612
A;Accession: A41914
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-208 <NAG>
A;Cross-references: UNIPROT:Q09118; GB:M93012
R;Hooper, K.P.; Eidels, L.
Biochem. Biophys. Res. Commun. 220, 675-680, 1996
A;Title: Glutamic acid 141 of the diptheria toxin receptor (HB-EGF precursor) is critica
A;Reference number: JC4687; MUID:96183008; PMID:8607824
A;Accession: JC4687
A;Molecule type: mRNA
A;Residues: 1-208 <HOO>
C;Comment: This precursor is a transmembrane glycoprotein which is cleaved at th e cell
C;Superfamily: heparin-binding EGF-like growth factor; EGF homology
C;Keywords: receptor; transmembrane protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;63-150/Product: diptheria toxin receptor precursor #status predicted <MAT>
F;108-143/Domain: EGF homology <EGF>
F;160-184/Domain: transmembrane #status predicted <TRN>

Query Match 16.2%; Score 211.5; DB 1; Length 208;
Best Local Similarity 36.7%; Pred. No. 2.3e-09;
Matches 55; Conservative 24; Mismatches 58; Indels 13; Gaps 4;

Qy 100 DSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGEC 159
Db 70 DLLRVTLSSKPQALATPSK--EEHGKRKKKGKGLG------KKRDPCLRKYKDFCIHGEC 121

Qy 160 KYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVI 219
Db 122 KYVKELRAPSCICHPGYHGERCHGLSLPVENRLYTYDHTTILAVVAVVLSSVCL---LVI 178

Qy 220 TVQLRRQYVRK--YEGEAEERKKLRQENGN 247
Db 179 VGLLMFRYHRRGGYDVENEEKVKLGMTNSH 208

RESULT 6
JC1410
heparin-binding EGF-like growth factor precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: JC1410; JC4659
R;Abraham, J.A.; Damm, D.; Bajardi, A.; Miller, J.; Klagsbrun, M.; Ezekowitz, R.A.B.
Biochem. Biophys. Res. Commun. 190, 125-133, 1993
A;Title: Heparin-binding EGF-like growth factor: Characterization of rat and mouse cDNA
A;Reference number: JC1409; MUID:93135756; PMID:7678488
A;Accession: JC1410
A;Molecule type: mRNA
A;Residues: 1-208 <ABR>
A;Cross-references: UNIPROT:Q06186; GB:L07264; NID:g192999; PIDN:AAA37542.1; PID:g19300
R;Harding, P.A.; Brigstock, D.R.; Shen, L.; Crissman-Combs, M.A.; Besner, G.E.
Gene 169, 291-292, 1996
A;Title: Characterization of the gene encoding murine heparin-binding epidermal growth
A;Reference number: JC4659; MUID:96194822; PMID:8647467
A;Accession: JC4659
A;Molecule type: DNA
A;Residues: 1-208 <HAR>
A;Cross-references: GB:U39189
C;Comment: This factor is a member of the epidermal growth factor family. Its binding a
C;Genetics:
A;Gene: mHB-EGF
A;Introns: 16/1; 74/1; 133/2; 185/2
C;Superfamily: heparin-binding EGF-like growth factor; EGF homology
C;Keywords: growth factor; heparin binding; transmembrane protein
F;1-23/Domain: signal sequence #status predicted <SIG>
F;24-208/Product: heparin-binding EGF-like growth factor #status predicted <MAT>
F;108-143/Domain: EGF homology <EGF>
F;161-184/Domain: transmembrane #status predicted <TMM>

Query Match 14.2%; Score 185.5; DB 1; Length 208;
Best Local Similarity 32.5%; Pred. No. 2.5e-07;
Matches 39; Conservative 28; Mismatches 48; Indels 5; Gaps 2;

Qy 124 PKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGE 183
Db 86 PSKERNGKKKKKGKGLGKKRDPCLRKYKDYCIHGECRYLQEFRTPSCKCLPGYHGHRCHG 145

Qy 184 KSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRK--YEGEAEERKKL 241
Db 146 LTLPVENPLYTYDHTTVLAVVAVVLSSVCL---LVIVGLLMFRYHRRGGYDLESEEKVKL 202

RESULT 7
JC1409
heparin-binding EGF-like growth factor precursor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: JC1409
R;Abraham, J.A.; Damm, D.; Bajardi, A.; Miller, J.; Klagsbrun, M.; Ezekowitz, R.A.B.
Biochem. Biophys. Res. Commun. 190, 125-133, 1993
A;Title: Heparin-binding EGF-like growth factor: Characterization of rat and mouse cDNA
A;Reference number: JC1409; MUID:93135756; PMID:7678488
A;Accession: JC1409
A;Molecule type: mRNA
A;Residues: 1-208 <ABR>
A;Cross-references: UNIPROT:Q06175; GB:L05489; NID:g204289; PIDN:AAA81780.1; PID:g20429
C;Superfamily: heparin-binding EGF-like growth factor; EGF homology
C;Keywords: growth factor; heparin binding; transmembrane protein
F;1-23/Domain: signal sequence #status predicted <SIG>

```
F;24-208/Product: heparin-binding EGF-like growth factor #status predicted <MAT>
F;108-143/Domain: EGF homology <EGF>
F;161-184/Domain: transmembrane #status predicted <TMM>

  Query Match             14.1%;  Score 184.5;  DB 1;  Length 208;
  Best Local Similarity   33.3%;  Pred. No. 2.9e-07;
  Matches   40;  Conservative   28;  Mismatches   47;  Indels    5;  Gaps    2;

Qy        124 PKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGE 183
              | ::| | |  | :  ||::||  :::::|||||:|:: |   :| |   | |:||
Db         86 PGKEKNGKKKRKGKGLGKKRDPCLKKYKDYCIHGECRYLKELRIPSCHCLPGYHGQRCHG 145

Qy        184 KSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRK--YEGEAEERKKL 241
              ::   : : :      || :|  :|:| |    ||   | :| |:  |: |:||: ||
Db        146 LTLPVENPLYTYDHTTVLAVVAVVLSSVCL---LVIVGLLMFRYHRRGGYDLESEEKVKL 202


RESULT 8
S27162
heparin-binding epidermal growth factor - pig (fragment)
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 09-Jul-2004
C;Accession: S27162; S25158
R;Vaughan, T.J.; Pascall, J.C.; Brown, K.D.
Biochem. J. 287, 681-684, 1992
A;Title: Tissue distribution of mRNA for heparin-binding epidermal growth factor.
A;Reference number: S27162; MUID:93075016; PMID:1445231
A;Accession: S27162
A;Molecule type: mRNA
A;Residues: 1-84 <VAU>
A;Cross-references: UNIPROT:Q01580; EMBL:X67295; NID:g1969; PIDN:CAA47709.1; PID:g1970
C;Superfamily: heparin-binding EGF-like growth factor; EGF homology
C;Keywords: growth factor; heparin binding
F;10-45/Domain: EGF homology <EGF>

  Query Match             12.4%;  Score 162;  DB 2;  Length 84;
  Best Local Similarity   38.2%;  Pred. No. 6.1e-06;
  Matches   29;  Conservative   17;  Mismatches   30;  Indels    0;  Gaps    0;

Qy        141 KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIA 200
              ||::||  ::::|||||||||:: | | :| |   | ||||   |:   : : :
Db          5 KKRDPCLRKYKDFCIHGECKYVKELRAPSCICHPGYHGERCHGLSLPVKNRLYTYDHTTI 64

Qy        201 LAAIAAFMSAVILTAV 216
              || :|  :|:| |  :
Db         65 LAVVAVVLSSVCLLVI 80


RESULT 9
G01639
transmembrane protein - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 09-Jul-2004
C;Accession: G01639
R;Eib, D.W.
submitted to the EMBL Data Library, January 1995
A;Reference number: G07997
A;Accession: G01639
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-380 <EIB>
A;Cross-references: UNIPROT:Q13086; EMBL:U19878; NID:g755465; PIDN:AAA64622.1; PID:g7554
F;275-310/Domain: EGF homology <EGF>

  Query Match             11.0%;  Score 144.5;  DB 2;  Length 380;
  Best Local Similarity   21.7%;  Pred. No. 0.00076;
  Matches   54;  Conservative   45;  Mismatches   95;  Indels   55;  Gaps    9;

Qy         24 AAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEM----------------------SS 59
              | |   :|  ||  |   |:    | ||  :                        |
Db        139 ARGPCYSDNGSGSGE---GEEEGSGAEVHRKHSKCGPCKYKAECDEDAENVGCVCNIDCS 195

Qy         60 GSEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDS-VRVEQV-VKPPQNKTES 117
              |   :||     ||              |:|    | :: : | :: ||: ::   : |::
Db        196 GYSFNPVCASDGSS-------------YNN----PCFVREASCIKQEQIDIRHLGHCTDT 238

Qy        118 ENTS-----DKPKRKKKGGKNGKNRRNR---KKKNPCNAEFQNFCIHGECKYIEHLEAVT 169
              ::||     |   : :   |: ::|          ||       :||||:|::|  |   :
Db        239 DDTSLLGKKDDGLQYRPDVKDASDQREDVYIGNHMPCPENLNGYCIHGKCEFIYLLRRAS 298

Qy        170 CKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVR 229
              |:|:  | |: | ||:  :   :  |  |:    ||| : || :  :  | : : |:  :
Db        299 CRCESGYTGQHC-EKTDFSILYVVPSRQKLTHVLIAAIIGAVQIAIIVAIVMCITRKCPK 357

Qy        230 KYEGEAEER 238
                |  :::
Db        358 NNRGRRQKQ 366


RESULT 10
T46914
hypothetical protein DKFZp564L1878.1 - human
C;Species: Homo sapiens (man)
C;Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 09-Jul-2004
C;Accession: T46914
R;Ottenwaelder, B.; Obermaier, B.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, February 2000
A;Reference number: Z24136
A;Accession: T46914
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-346 <AAA>
A;Cross-references: UNIPROT:Q9NSS5; EMBL:AL157430
A;Experimental source: fetal brain; clone DKFZp564L1878
C;Genetics:
A;Note: DKFZp564L1878.1

  Query Match             10.9%;  Score 143;  DB 2;  Length 346;
  Best Local Similarity   20.3%;  Pred. No. 0.0009;
  Matches   47;  Conservative   46;  Mismatches  104;  Indels   34;  Gaps    8;

Qy         15 LLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSS------GSEISPVSE 68
              :|::  |  |       |  ||     |||   :|   ||: | |:      |:|    :|
Db        127 ILVVSEGSCA-----TDAGSG-----SGDGVHEGSGETSQKETSTCDICQFGAECDEDAE 176

Qy         69 ----------MPSSSEPSSGADYDYSEEYDNEPQI--PGYIVDDSVRVEQVVKPPQNKTE 116
                        ::  |   :|    : ||| ||       : : |  : :    | |
Db        177 DVWCVCNIDCSQTNFNPLCASD---GKSYDNACQIKEASCQKQEKIEVMSLGRCQDNTTT 233

Qy        117 SENTSDKPKRKKKGGKNGKNRRNRKKKN--PCNAEFQNFCIHGECKYIEHLEAVTCKCQQ 174
              :  : |   :   :|        :::  ||   :  ||:||:|::  :::  :|:|
Db        234 TTKSEDGHYARTDYAENANKLEESAREHHIPCPEHYNGFCMHGKCEHSINMQEPSCRCDA 293

Qy        175 EYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRR 225
              | |: | :|      ::   : :     |||  :  : :  : |: : : |
Db        294 GYTGQHCEKKDYSVLYVVPGPV-RFQYVLIAAVIGTIQIAVICVVVLCITR 343


RESULT 11
S58117
EGF-like growth factor, precursor - green monkey
C;Species: Cercopithecus aethiops (green monkey, grivet)
C;Date: 13-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 09-Jul-2004
C;Accession: S58117
R;Loukianov, E.V.; Loukianov, T.I.; Wiedlocha, A.; Olsnes, S.
submitted to the EMBL Data Library, July 1995
A;Description: Short form of heparin binding EGF-like growth factor.
A;Reference number: S58117
A;Accession: S58117
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-141 <LOU>
A;Cross-references: UNIPROT:Q28218; EMBL:X89728; NID:g1052560; PIDN:CAA61880.1; PID:g105
C;Superfamily: heparin-binding EGF-like growth factor; EGF homology

  Query Match             10.5%;  Score 137;  DB 2;  Length 141;
```

```
  Best Local Similarity   42.3%;  Pred. No. 0.00097;
  Matches   30;  Conservative   12;  Mismatches   21;  Indels    8;  Gaps    2;

Qy        100 DSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGEC 159
              | :||    ||     | |:   :  ||||||   |       ||::||  ::::|||||||
Db         70 DLLRVTLSSKPQALATPSK--EEHGKRKKKGKGLG------KKRDPCLRKYKDFCIHGEC 121

Qy        160 KYIEHLEAVTC 170
              ||:: | | :|
Db        122 KYVKELRAPSC 132


RESULT 12
A37408
betacellulin precursor - mouse
C;Species: Mus musculus (house mouse)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Jul-2004
C;Accession: A37408
R;Shing, Y.; Christofori, G.; Hanahan, D.; Ono, Y.; Sasada, R.; Igarashi, K.; Folkman, J
Science 259, 1604-1607, 1993
A;Title: Betacellulin: a mitogen from pancreatic beta cell tumors.
A;Reference number: A37408; MUID:93206093; PMID:8456283
A;Accession: A37408
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-177 <SHI>
A;Cross-references: UNIPROT:Q05928; GB:L08394; NID:g293852; PIDN:AAA40511.1; PID:g293853
C;Keywords: transmembrane protein
F;69-104/Domain: EGF homology <EGF>

  Query Match             9.1%;  Score 119.5;  DB 2;  Length 177;
  Best Local Similarity   25.0%;  Pred. No. 0.029;
  Matches   29;  Conservative   25;  Mismatches   53;  Indels    9;  Gaps    4;

Qy        131 GKN--GKNRRNRKKK--NPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSM 186
              |:|  |   | : |   : |  :::::|||| |:::   :  :| |:: ||| ||    :
Db         50 GENCTGTTPRQKVKTHFSRCPKQYKHYCIHGRCRFVVDEQTPSCICEKGYFGARCERVDL 109

Qy        187 KTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITV--QLRRQYVRKYEGEAEERKK 240
                :     :| :  :   |   |:  : | |    ||:   :| | : |   |
Db        110 -+-FYLQQDRGQILVVCLIVVMVVFIILVIGVCTCCHPLRKHRKKKKEEKMETLDK 162


RESULT 13
JC5702
ErbB kinase activator alpha2a, brain and thymus - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 25-Nov-1997 #sequence_revision 25-Nov-1997 #text_change 09-Jul-2004
C;Accession: JC5702; PC4417
R;Higashiyama, S.; Horikawa, M.; Yamada, K.; Ichino, N.; Nakano, N.; Nakagawa, T.; Miyag
J. Biochem. 122, 675-680, 1997
A;Title: A novel brain-derived member of the epidermal growth factor family that interac
A;Reference number: JC5700; MUID:98006324; PMID:9348101
A;Accession: JC5702
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-860 <HIG>
A;Cross-references: UNIPROT:O35569; DDBJ:D89996; NID:g2605631; PIDN:BAA23345.1; PID:g260
A;Experimental source: PC-12 cell
A;Accession: PC4417
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 'F',212-213,223-860 <HI2>
A;Cross-references: DDBJ:AB001576; NID:g2605478; PIDN:BAA23348.1; PID:g2605479
A;Experimental source: PC-12 cell
C;Comment: This protein is a member of the epidermal growth factor family. It is functic
ating the differentiation of MDA-MB-453 cells.
C;Superfamily: human ErbB kinase activator alpha, brain and thymus; EGF homology; immunc
C;Keywords: glycoprotein
F;274-327/Domain: Ig-like #status predicted <IGL>
F;361-397/Domain: EGF homology <EGF>
F;422-444/Domain: hydrophobic #status predicted <HYD>
F;163,294,467/Binding site: carbohydrate (Asn) (covalent) #status predicted

  Query Match             8.9%;  Score 117;  DB 2;  Length 860;
  Best Local Similarity   25.4%;  Pred. No. 0.26;
  Matches   30;  Conservative   22;  Mismatches   50;  Indels   16;  Gaps    4;

Qy        146 CNAEFQNFCIH-GECKYIEHLEAVTCKCQQEYFGERCGEK----------SMKTHSMIDS 194
              ||   :::|:: | | ||| :  ::|||   :||:|| ||            |   :
Db        361 CNETAKSYCVNGGVCYYIEGINQLSCKCPNGFFGQRCLEKLPLRLYMPDPKQKAEELYQK 420

Qy        195 SLSKIALAAIAAFMSAVILTAVAVITVQLRRQ---YVRKYEGEAEERKKLRQENGNVH 249
              : |   :|  :  ::        | : |||   ::|:    | : : |   ||  |
Db        421 RVLTITGICVALLVVGIVCVVAYCKTKKQRRQMHHHLRQNMCPAHQNRSL--ANGPSH 476


RESULT 14
B71618
merozoite surface antigen MSP-4 (EGF domain) PFB0310c - malaria parasite (Plasmodium fa
C;Species: Plasmodium falciparum
C;Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 09-Jul-2004
C;Accession: B71618
R;Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.
.; Pertea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O
Science 282, 1126-1132, 1998
A;Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A;Reference number: A71600; MUID:99021743; PMID:9804551
A;Accession: B71618
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-272 <GAR>
A;Cross-references: UNIPROT:Q8T7U3; UNIPROT:Q8T7U7; UNIPROT:Q8T7U6; UNIPROT:Q8T7T9; UNI
C
A;Experimental source: clone 3D7
C;Genetics:
A;Gene: PFB0310c

  Query Match             8.8%;  Score 115;  DB 2;  Length 272;
  Best Local Similarity   21.9%;  Pred. No. 0.1;
  Matches   55;  Conservative   43;  Mismatches  101;  Indels   52;  Gaps    9;

Qy          5 LLPPAPVVLSLLILGSGH-YAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEI 63
              ::|    :|:: |||       |:  ::|  |     :  : :|  |     | | |
Db         31 IVPENGRMLNMRILGEEKPNVDGVSTSNTPGGNESSSASPNLSDAAEKKDEKEASEQGEE 90

Qy         64 SPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDK 123
              |   |  :| | ::| |    |:  || :  |    | |: : : | |:   ||:   ||
Db         91 SHKKE--NSQESANGKDDVKEEKKTNEKKDDG--KTDKVQEKVLEKSPK---ESQMVDDK 143

Qy        124 ------------PKRKKKGGKNGKNRRNRKKK--------------------NPCNAEF 150
                         |     ||  |:   : : :                       |  : |
Db        144 KKTEAIPKKVVQPSSSNSGGHVGEEEDHNEGEGEHEEEEEHEEDDDDEDDDTYNKDDLED 203

Qy        151 QNFCIHGE--------CKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALA 202
              :: | |           |:|: :   | |||:: |  |  | : ::  |:  |||: | :
Db        204 EDLCKHNNGGCGDDKLCEYVGN-RRVKCKCKEGYKLE--GIECVELLSLASSSLNLIFNS 260

Qy        203 AIAAFMSAVIL 213
              |  |:  :::
Db        261 FITIFVVILLI 271


RESULT 15
EGHU
epidermal growth factor precursor [validated] - human
N;Alternate names: urogastrone precursor
C;Species: Homo sapiens (man)
C;Date: 30-Nov-1980 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: A25531; A01388; A33517; A29721; S45282; S45283
R;Bell, G.I.; Fong, N.M.; Stempien, M.M.; Wormsted, M.A.; Caput, D.; Ku, L.; Urdea, M.S
Nucleic Acids Res. 14, 8427-8446, 1986
A;Title: Human epidermal growth factor precursor: cDNA sequence, expression in vitro an
A;Reference number: A25531; MUID:87066721; PMID:3491360
A;Accession: A25531
A;Molecule type: mRNA
```

A;Residues: 1-1207 <BEL>
A;Cross-references: UNIPROT:P01133; EMBL:X04571; NID:g31120; PIDN:CAA28240.1; PID:g31121
A;Note: 708-Met was also found
A;Note: intron positions were also determined
R;Gregory, H.; Preston, B.M.
Int. J. Pept. Protein Res. 9, 107-118, 1977
A;Title: The primary structure of human urogastrone.
A;Reference number: A01388; MUID:77117897; PMID:300079
A;Accession: A01388
A;Molecule type: protein
A;Residues: 971-1023 <GRE>
A;Note: some of the molecules lack Arg-1023
R;Furuya, M.; Akashi, S.; Hirayama, K.
Biochem. Biophys. Res. Commun. 163, 1100-1106, 1989
A;Title: The primary structure of human EGF produced by genetic engineering, studied by
A;Reference number: A33517; MUID:89391964; PMID:2789514
A;Accession: A33517
A;Molecule type: protein
A;Residues: 971-1023 <FUR>
R;Tsukumo, K.; Nakamura, H.; Sakamoto, S.
Biochem. Biophys. Res. Commun. 145, 126-133, 1987
A;Title: Purification and characterization of high molecular weight human epidermal grow
A;Reference number: A29721; MUID:87241488; PMID:3297054
A;Accession: A29721
A;Molecule type: protein
A;Residues: 829-834,'X',836-839,'X',841-845,'X',847-848 <TSU>
A;Note: this is the amino-terminal sequence of a high molecular weight form of EGF, isol
R;Svoboda, M.; Bauhofer, A.; Schwind, P.; Bade, E.; Rasched, I.; Przybylski, M.
Biochim. Biophys. Acta 1206, 35-41, 1994
A;Title: Structural characterization and biological activity of recombinant human epider
A;Reference number: S45282; MUID:94242778; PMID:8186248
A;Accession: S45282
A;Molecule type: protein
A;Residues: 'M',971-1023 <SVO>
A;Note: expressed recombinant protein
A;Accession: S45283
A;Molecule type: protein
A;Residues: 'MKKYP',970-1023 <SV2>
A;Note: expressed recombinant protein
C;Comment: Epidermal growth factor (EGF) stimulates the proliferation and differentiatic
gastrointestinal cell proliferation.
C;Comment: EGF is released in the pancreas, small intestine, mammary gland, and (in some
C;Comment: The EGF precursor is found in kidney as a receptor-like membrane-bound protei
C;Genetics:
A;Gene: GDB:EGF
A;Cross-references: GDB:119105; OMIM:131530
A;Map position: 4q25-4q25
A;Introns: 43/1; 109/3; 170/2; 246/2; 314/1; 356/3; 397/1; 438/1; 480/1; 525/3; 575/2; 6
C;Superfamily: epidermal growth factor precursor; EGF homology; LDL receptor YWTD-contai
C;Keywords: duplication; growth factor; tandem repeat; transmembrane protein
F;1-22/Domain: signal sequence #status predicted <SIG>
F;23-1207/Product: epidermal growth factor proprotein, membrane-bound form #status predi
F;23-1032/Domain: extracellular #status predicted <EXT>
F;43-479/Region: EGF precursor long repeat <LR1>
F;46-85/Domain: LDL receptor YWTD-containing repeat homology <YW01>
F;86-127/Domain: LDL receptor YWTD-containing repeat homology <YW02>
F;128-169/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F;170-211/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F;212-256/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F;257-301/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F;318-354/Domain: EGF homology <EG1>
F;360-395/Domain: EGF homology <EG2>
F;401-436/Domain: EGF homology <EG3>
F;439-476/Domain: EGF homology <EG4>
F;480-954/Region: EGF precursor long repeat <LR2>
F;483-523/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F;524-566/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F;567-609/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F;610-653/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F;654-694/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F;695-737/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F;745-780/Domain: EGF homology <EG5>
F;835-868/Domain: EGF homology <EG6>
F;874-910/Domain: EGF homology <EG7>
F;916-951/Domain: EGF homology <EG8>
F;971-1023/Product: epidermal growth factor #status experimental <EGF>
F;976-1012/Domain: EGF homology <EG9>
F;1033-1057/Domain: transmembrane #status predicted <TMM>
F;1058-1207/Domain: intracellular #status predicted <INT>
F;318-330,325-339,341-354,360-371,367-380,382-395,401-412,408-421,423-436,439-451,447-4
fide bonds: #status predicted
F;976-990,984-1001,1003-1012/Disulfide bonds: #status experimental

Query Match 8.8%; Score 114.5; DB 1; Length 1207;
Best Local Similarity 22.5%; Pred. No. 0.6;
Matches 50; Conservative 22; Mismatches 95; Indels 55; Gaps 9;

```
Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              |  |: |||    |  |   | |       :: | |             ||        |   |
Db        876 MGVPVCPPAS---SKCINTEGGYVC--RCSEGYQG----------DGIHCLDIDECQLG 919

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
                :   |  | :    |    |:          || |  ||        || :  | ::
Db        920 --VHSCGENASCTNTEGG----YTCMCAGRLSEPGLICPDST-------PPPHLREDDH- 965

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIH-GECKYIEHLEAVTCKCQQEYFGE 179
                           :   |   : |      :|:| | | ||| |:   | |   | ||
Db        966 --------------HYSVRNSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGE 1010

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITV 221
              ||  : :|   :  :  :              ||: || |: :
Db       1011 RCQYRDLKWWELRHAGHGQ---------QQKVIVVAVCVVVL 1043
```

Search completed: February 26, 2005, 16:45:00
Job time : 42 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: February 26, 2005, 15:50:20 ; Search time 174 Seconds
(without alignments)
741.632 Million cell updates/sec

Title: US-10-774-076-1
Perfect score: 1308
Sequence: 1 MRAPLLPPAPVVLSLLILGS.....GEAEERKKLRQENGNVHAIA 252

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1308	100.0	252	1	AMPR_HUMAN	P15514 homo sapien
2	1059.5	81.0	247	2	Q9BDH4	Q9bdh4 sus scrofa
3	941	71.9	216	2	Q9BDH3	Q9bdh3 sus scrofa
4	918	70.2	248	1	AMPR_MOUSE	P31955 mus musculu
5	877	67.0	248	2	Q62561	Q62561 mesocricetu
6	874.5	66.9	243	1	AMPR_RAT	P24338 rattus norv
7	607.5	46.4	165	2	Q8K3E3	Q8k3e3 mesocricetu
8	435.5	33.3	210	2	Q645M5	Q645m5 gallus gall
9	348	26.6	77	2	P79199	P79199 ovis aries
10	224.5	17.2	239	2	Q6NY79	Q6ny79 brachydanio
11	211.5	16.2	208	1	HBGF_CERAE	Q09118 cercopithec
12	211.5	16.2	208	1	HBGF_HUMAN	Q99075 homo sapien
13	208.5	15.9	208	1	HBGF_PIG	Q01580 sus scrofa
14	198.5	15.2	212	2	Q9W7C5	Q9w7c5 gallus gall
15	195	14.9	132	2	Q9BG62	Q9bg62 oryctolagus
16	187	14.3	188	2	Q9R0C7	Q9r0c7 cricetulus
17	185.5	14.2	208	1	HBGF_MOUSE	Q06186 mus musculu
18	184.5	14.1	208	1	HBGF_RAT	Q06175 rattus norv
19	183	14.0	105	2	Q99MN7	Q99mn7 mesocricetu
20	150	11.5	374	2	Q9QYM9	Q9qym9 m tmeff2 pr
21	145	11.1	368	2	Q9P2Y9	Q9p2y9 homo sapien
22	145	11.1	374	2	Q8N2R5	Q8n2r5 homo sapien
23	145	11.1	374	2	Q9UIK5	Q9uik5 homo sapien
24	144.5	11.0	380	2	Q13086	Q13086 homo sapien
25	143.5	11.0	341	2	Q8N3T8	Q8n3t8 homo sapien
26	143.5	11.0	380	2	Q8IYR6	Q8iyr6 homo sapien
27	143	10.9	346	2	Q9NSS5	Q9nss5 homo sapien
28	142	10.9	373	2	Q9QYV1	Q9qyv1 rattus norv
29	137	10.5	141	2	Q28218	Q28218 cercopithec
30	136	10.4	354	2	Q9JJS1	Q9jjs1 mus musculu
31	134	10.2	247	2	Q8C536	Q8c536 mus musculu
32	133	10.2	256	2	Q8BRP7	Q8brp7 mus musculu
33	133	10.2	372	2	Q6PFE7	Q6pfe7 mus musculu
34	128	9.8	172	2	Q91590	Q91590 xenopus lae
35	124	9.5	272	2	Q8T7T2	Q8t7t2 plasmodium
36	124	9.5	272	2	Q8T7T3	Q8t7t3 plasmodium
37	124	9.5	272	2	Q8T7T8	Q8t7t8 plasmodium
38	123.5	9.4	177	2	Q9JJM4	Q9jjm4 rattus norv
39	123	9.4	272	2	Q8ITM3	Q8itm3 plasmodium
40	120	9.2	272	2	Q8I0F0	Q8i0f0 plasmodium
41	120	9.2	272	2	Q8I0I8	Q8i0i8 plasmodium
42	120	9.2	272	2	Q8ITL6	Q8itl6 plasmodium
43	120	9.2	272	2	Q8ITL8	Q8itl8 plasmodium
44	120	9.2	272	2	Q70ID2	Q70id2 plasmodium
45	120	9.2	272	2	Q70IE8	Q70ie8 plasmodium

ALIGNMENTS

RESULT 1
AMPR_HUMAN
ID AMPR_HUMAN STANDARD; PRT; 252 AA.
AC P15514;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Amphiregulin precursor (AR) (Colorectum cell-derived growth factor)
DE (CRDGF).
GN Name=AREG;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220581; PubMed=2325643;
RA Plowman G.D., Green J.M., McDonald V.L., Neubauer M.G., Disteche C.M.,
RA Todaro G.J., Shoyab M.;
RT "The amphiregulin gene encodes a novel epidermal growth factor-related
RT protein with tumor-inhibitory activity.";
RL Mol. Cell. Biol. 10:1969-1981(1990).
RN [2]
RP SEQUENCE FROM N.A., AND VARIANTS VAL-80 AND CYS-81.
RA Rieder M.J., Livingston R.J., Daniels M.R., Chung M.-W.,
RA Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RT "NIEHS-SNPs, environmental genome project, NIEHS ES15478, Department
RT of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
RL Submitted (OCT-2003) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```
RN   [4]
RP   SEQUENCE OF 101-184.
RX   MEDLINE=89162036; PubMed=2466334;
RA   Shoyab M., Plowman G.D., McDonald V.L., Bradley J.G., Todaro G.J.;
RT   "Structure and function of human amphiregulin: a member of the
RT   epidermal growth factor family.";
RL   Science 243:1074-1076(1989).
RN   [5]
RP   SEQUENCE OF 101-131.
RX   MEDLINE=88320474; PubMed=3413110;
RA   Shoyab M., McDonald V.L., Bradley J.G., Todaro G.J.;
RT   "Amphiregulin: a bifunctional growth-modulating glycoprotein produced
RT   by the phorbol 12-myristate 13-acetate-treated human breast
RT   adenocarcinoma cell line MCF-7.";
RL   Proc. Natl. Acad. Sci. U.S.A. 85:6528-6532(1988).
RN   [6]
RP   SEQUENCE OF 101-126.
RX   MEDLINE=91203875; PubMed=2017164;
RA   Cook P.W., Mattox P.A., Keeble W.W., Pittelkow M.R., Plowman G.D.,
RA   Shoyab M., Adelman J.P., Shipley G.D.;
RT   "A heparin sulfate-regulated human keratinocyte autocrine factor is
RT   similar or identical to amphiregulin.";
RL   Mol. Cell. Biol. 11:2547-2557(1991).
RN   [7]
RP   SEQUENCE OF 101-122.
RX   MEDLINE=93090386; PubMed=1333777;
RA   Culouscou J.-M., Remacle-Bonnet M., Carlton G.W., Plowman G.D.,
RA   Shoyab M.;
RT   "Colorectum cell-derived growth factor (CRDGF) is homologous to
RT   amphiregulin, a member of the epidermal growth factor family.";
RL   Growth Factors 7:195-205(1992).
CC   -!- FUNCTION: Bifunctional growth-modulating glycoprotein. Inhibits
CC       growth of several human carcinoma cells in culture and stimulates
CC       proliferation of human fibroblasts and certain other tumor cells.
CC   -!- INDUCTION: By phorbol 12-myristate 13-acetate (PMA).
CC   -!- MISCELLANEOUS: AR is a protein containing cysteines in disulfide
CC       linkage(s) that are essential for its biological activity. AR may
CC       contain oligosaccharides and/or lipid moieties that are not
CC       obligatory for the biological activity.
CC   -!- SIMILARITY: Belongs to the amphiregulin family.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M30704; AAA51781.1; -.
DR   EMBL; M30702; AAA51773.1; -.
DR   EMBL; M30698; AAA51773.1; JOINED.
DR   EMBL; M30699; AAA51773.1; JOINED.
DR   EMBL; M30700; AAA51773.1; JOINED.
DR   EMBL; M30701; AAA51773.1; JOINED.
DR   EMBL; AY442340; AAR05438.1; -.
DR   EMBL; BC009799; AAH09799.1; -.
DR   PIR; A34702; A34702.
DR   HSSP; Q99075; 1XDT.
DR   Genew; HGNC:651; AREG.
DR   H-InvDB; HIX0004290; -.
DR   MIM; 104640; -.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0008083; F:growth factor activity; TAS.
DR   GO; GO:0008283; P:cell proliferation; TAS.
DR   GO; GO:0007267; P:cell-cell signaling; TAS.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 1.
KW   Cytokine; Direct protein sequencing; EGF-like domain; Glycoprotein;
KW   Growth factor; Polymorphism; Signal; Transmembrane.
FT   SIGNAL       1     19       Potential.
FT   PROPEP       20    100
FT   CHAIN        101   184      Amphiregulin.
FT   PROPEP       185   252
FT   DOMAIN       142   182      EGF-like.
FT   TRANSMEM     199   221      Potential.
FT   DISULFID     146   159      By similarity.
FT   DISULFID     154   170      By similarity.
FT   DISULFID     172   181      By similarity.
FT   CARBOHYD     30    30       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     113   113      N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD     119   119      N-linked (GlcNAc. . .).
FT   VARIANT      80    80       D -> V.
FT                               /FTId=VAR_018918.
FT   VARIANT      81    81       Y -> C.
FT                               /FTId=VAR_018919.
SQ   SEQUENCE   252 AA;  27895 MW;  68C1FA6BCD4D2BED CRC64;

  Query Match           100.0%;  Score 1308;  DB 1;  Length 252;
  Best Local Similarity 100.0%;  Pred. No. 5.4e-93;
  Matches  252;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKK 240

Qy        241 LRQENGNVHAIA 252
              ||||||||||||
Db        241 LRQENGNVHAIA 252


RESULT 2
Q9BDH4
ID   Q9BDH4       PRELIMINARY;      PRT;   247 AA.
AC   Q9BDH4;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Amphiregulin long form.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim J.G., Vallet J.L., Rohrer G.A., Christenson R.K.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
DR   EMBL; AY028310; AAK27728.1; -.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain.
FT   CHAIN        95    178      amphiregulin long form.
```

```
SQ   SEQUENCE   247 AA;  27336 MW;  6D1FA68DFE4B9957 CRC64;

  Query Match          81.0%;  Score 1059.5;  DB 2;  Length 247;
  Best Local Similarity   80.6%;  Pred. No. 7.9e-74;
  Matches  204;  Conservative   19;  Mismatches   23;  Indels    7;  Gaps    2;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              |||||||||||||||||| | |||||||:| | ||| ||||||| |: ||||||||||||
Db          1 MRAPLLPPAPVVLSLLILCSAHYAAGLDVNGTSSGKGEPFSGDHGAEAFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||  | ||||      ||:||||:||||||| | |||||||||||||| :|:||||||
Db         61 SEAPPASEMP------SGSDYDYAEEYDNEPHISGYIVDDSVRVEQVVKPKRNRTESENT 114

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              |||||||||||:|||||||||||||:||||||||||:|||||||||||||| |:|||||
Db        115 SDKPKRKKKGGKSGKNRRNRKKKNPCDAEFQNFCIHGDCKYIEHLEAVTCKCYQDYFGER 174

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVA-VITVQLRRQYVRKYEGEAEERK 239
              |||||||||:|: | |||||||||||||:||:  ||:| |||: ||::| |:||| |||||
Db        175 CGEKSMKTHTMVHSDLSKIALAAIAAFVSAMSFTAIAVVITIYLRKRYFREYEGAAEERK 234

Qy        240 KLRQENGNVHAIA 252
              |||||| | ||||
Db        235 KLRQENANAHAIA 247


RESULT 3
Q9BDH3
ID   Q9BDH3      PRELIMINARY;      PRT;   216 AA.
AC   Q9BDH3;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Amphiregulin short form.
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Kim J.G., Vallet J.L., Rohrer G.A., Christenson R.K.;
RL   Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
DR   EMBL; AY028311; AAK27729.1; -.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain.
FT   CHAIN       95    178       amphiregulin short form.
SQ   SEQUENCE   216 AA;  23571 MW;  5DA65FF187ED89DD CRC64;

  Query Match          71.9%;  Score 941;  DB 2;  Length 216;
  Best Local Similarity   81.7%;  Pred. No. 9.3e-65;
  Matches  179;  Conservative   16;  Mismatches   18;  Indels    6;  Gaps    1;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              |||||||||||||||||| | |||||||:| | ||| ||||||| |: ||||||||||||
Db          1 MRAPLLPPAPVVLSLLILCSAHYAAGLDVNGTSSGKGEPFSGDHGAEAFEVTSRSEMSSG 60

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||  | ||||      ||:||||:||||||| | |||||||||||||| :|:||||||
Db         61 SEAPPASEMP------SGSDYDYAEEYDNEPHISGYIVDDSVRVEQVVKPKRNRTESENT 114

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              |||||||||||:|||||||||||||:||||||||||:|||||||||||||| |:|||||
Db        115 SDKPKRKKKGGKSGKNRRNRKKKNPCDAEFQNFCIHGDCKYIEHLEAVTCKCYQDYFGER 174

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVI 219
              |||||||||:|: | |||||||||||||:||:  ||:||:
Db        175 CGEKSMKTHTMVHSDLSKIALAAIAAFVSAMSFTAIAVV 213


RESULT 4
AMPR_MOUSE
ID   AMPR_MOUSE      STANDARD;      PRT;   248 AA.
AC   P31955;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Amphiregulin precursor (AR) (Schwannoma-derived growth factor) (SDGF).
GN   Name=Areg; Synonyms=Sdgf;
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92287078; PubMed=1318038;
RA   Sonoda H., Yamaguchi T., Watanabe S.;
RT   "Androgen-responsive expression and mitogenic activity of schwannoma-
RT   derived growth factor on an androgen-dependent Shionogi mouse mammary
RT   carcinoma cell line.";
RL   Biochem. Biophys. Res. Commun. 185:103-109(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96174905; PubMed=8592515; DOI=10.1210/me.9.6.691;
RA   Das S.K., Chakraborty I., Paria B.C., Wang X.N., Plowman G.D.,
RA   Dey S.K.;
RT   "Amphiregulin is an implantation-specific and progesterone-regulated
RT   gene in the mouse uterus.";
RL   Mol. Endocrinol. 9:691-705(1995).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Cecum;
RX   MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA   Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA   Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA   Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA   Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA   Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA   Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA   Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA   Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA   Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA   Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA   Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA   Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA   Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA   Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA   Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA   Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA   Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA   Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA   Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA   Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA   Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA   Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA   Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA   Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA   Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA   Birney E., Hayashizaki Y.;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Mammary gland;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
```

```
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Autocrine growth factor as well as a mitogen for a broad
CC       range of target cells, among which astrocytes, Schwann cells and
CC       fibroblasts.
CC   -!- INDUCTION: Androgen-dependent.
CC   -!- SIMILARITY: Belongs to the amphiregulin family.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; D12648; BAA02169.1; -.
DR   EMBL; L41352; AAB00472.1; -.
DR   EMBL; AK018590; BAB31296.1; -.
DR   EMBL; BC009138; AAH09138.1; -.
DR   PIR; JH0612; JH0612.
DR   HSSP; Q99075; 1XDT.
DR   MGD; MGI:88068; Areg.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 1.
KW   Cytokine; EGF-like domain; Glycoprotein; Growth factor; Signal;
KW   Transmembrane.
FT   SIGNAL       1     26       Potential.
FT   PROPEP      27     99
FT   CHAIN      100    248       Amphiregulin.
FT   DOMAIN     135    175       EGF-like.
FT   TRANSMEM   192    215       Potential.
FT   DISULFID   139    152       By similarity.
FT   DISULFID   147    163       By similarity.
FT   DISULFID   165    174       By similarity.
FT   CARBOHYD   106    106       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD   241    241       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   248 AA;  27549 MW;  98C61A1B0E75A64E CRC64;

  Query Match             70.2%;  Score 918;  DB 1;  Length 248;
  Best Local Similarity   69.4%;  Pred. No. 6.6e-63;
  Matches  177;  Conservative   27;  Mismatches   41;  Indels   10;  Gaps    3;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              || |||| |  || ||:|||||||| |:|||  ||| |  |||||| |        |:| |
Db          1 MRTPLLPLARSVLLLLVLGSGHYAAALELNDPSSGKGESLSGDHSAGGL------ELSVG 54

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
               |:| :||||| || |:| |||||||||||||| |||:|||||||:|| :|||| | :
Db         55 REVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIIDDSVRVEQVIKPKKNKTEGEKS 113

Qy        121 SDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGER 180
              ::|||||||||||||| |||:|||||| |:||||||||||:|||:|| ||| | |:||||
Db        114 TEKPKRKKKGGKNGKGRRNKKKKNPCTAKFQNFCIHGECRYIENLEVVTCNCHQDYFGER 173

Qy        181 CGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVA---VITVQLRRQYVRKYEGEAEE 237
              |||||||||| |  |||||: |:  |:||:|| |:    |||| | ::| |:|||| ||
Db        174 CGEKSMKTHSEDDKDLSKIAVVAVTIFVSAIILAAIGIGIVITVHLWKRYFREYEGETEE 233

Qy        238 RKKLRQENGNVHAIA 252
              |::|||||| |||||
Db        234 RRRLRQENGTVHAIA 248


RESULT 5
Q62561
ID   Q62561      PRELIMINARY;      PRT;   248 AA.
AC   Q62561;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Schwannoma derived growth factor.
OS   Mesocricetus sp.
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10038;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97041064; PubMed=8975671;
RA   Gurskaya N.G., Shagin D.A., Luk'ianov K.A., Vagner L.L.,
RA   Shtutman M.S., Musatkin E.A., Moinova E.V., Tatosian A.G.,
RA   Luk'ianov S.A., Sverdlov E.D.;
RT   "Cloning cDNA for the ha-SDGF gene from a Syrian hamster cell line
RT   with increased metastatic potential using subtractive hybridization.";
RL   Bioorg. Khim. 22:425-431(1996).
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
DR   EMBL; Z66535; CAA91439.1; -.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain.
SQ   SEQUENCE   248 AA;  27381 MW;  CD920DD40B3D1063 CRC64;

  Query Match             67.0%;  Score 877;  DB 2;  Length 248;
  Best Local Similarity   69.4%;  Pred. No. 9.6e-60;
  Matches  177;  Conservative   26;  Mismatches   42;  Indels   10;  Gaps    5;

Qy          1 MRAPLLPPA-PVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSS 59
              || |||| | ||:  |||||||||||||||| | ||| : ||| |:|    |     |
Db          1 MRTPLLPRALPVLSLLLILGSGHYAAGLDLNGTTSGKSQAFSGHHNATELVV------SV 54

Qy         60 GSEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESEN 119
               | |:| :||||| || |:| |||||||||||||| ||::|||||||:|| :|| | :
Db         55 GREVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYVIDDSVRVEQVIKPKKNKAEGDK 113

Qy        120 TSDKPKRKKKGGKNGKNRRNRK-KKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFG 178
              |||||||:|||||:|| |||:| |||||: |||||||||||||||:||||||  |  :|||
Db        114 TSDKPKRRKKGGKSGKGRRNKKRKKNPCDTEFQNFCIHGECKYIENLEAVTCNCHLDYFG 173

Qy        179 ERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVA-VITVQLRRQYVRKYEGEAEE 237
              |||||||||||  |: |||||||||  |:||| | ||  |: | ||:::::|| ||||
Db        174 ERCGEKSMKTHSGDDNDLSKIALAAITVFVSAVTLIAVGIVVIVLLRKRHLKEYEEEAEE 233

Qy        238 RKKLRQENGNVHAIA 252
              ||||||| |||  ||
Db        234 RKKLRQETGNVPVIA 248


RESULT 6
AMPR_RAT
```

```
ID   AMPR_RAT       STANDARD;      PRT;   243 AA.
AC   P24338;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Amphiregulin precursor (AR) (Schwannoma-derived growth factor) (SDGF).
GN   Name=Areg; Synonyms=Sdgf;
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX   MEDLINE=91043095; PubMed=2234093; DOI=10.1038/348257a0;
RA   Kimura H., Fischer W.H., Schubert D.;
RT   "Structure, expression and function of a schwannoma-derived growth
RT   factor.";
RL   Nature 348:257-260(1990).
CC   -!- FUNCTION: Autocrine growth factor as well as a mitogen for a broad
CC       range of target cells, among which astrocytes, Schwann cells and
CC       fibroblasts.
CC   -!- SIMILARITY: Belongs to the amphiregulin family.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X55183; CAA38967.1; -.
DR   PIR; S13296; S13296.
DR   HSSP; Q99075; 1XDT.
DR   RGD; 2149; Areg.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; FALSE_NEG.
DR   PROSITE; PS50026; EGF_3; 1.
KW   Cytokine; Direct protein sequencing; EGF-like domain; Glycoprotein;
KW   Growth factor; Signal; Transmembrane.
FT   SIGNAL        1     24       Potential.
FT   PROPEP       25     96
FT   CHAIN        97    243       Amphiregulin.
FT   DOMAIN      133    173       EGF-like.
FT   TRANSMEM    190    213       Potential.
FT   DISULFID    137    150       By similarity.
FT   DISULFID    145    161       By similarity.
FT   DISULFID    163    172       By similarity.
FT   CARBOHYD     30     30       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    103    103       N-linked (GlcNAc. . .) (Potential).
FT   CARBOHYD    236    236       N-linked (GlcNAc. . .) (Potential).
SQ   SEQUENCE   243 AA;  26633 MW;  F59A8173D4921599 CRC64;

  Query Match             66.9%;  Score 874.5;  DB 1;  Length 243;
  Best Local Similarity   69.7%;  Pred. No. 1.5e-59;
  Matches  177;  Conservative   24;  Mismatches   40;  Indels   13;  Gaps    5;

Qy          1 MRAPLLPPAPVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSG 60
              || | |  |  |||||:||||||||||:|| | ||| || ||||||  |  |
Db          1 MRTPSLSLALSVLSLLVLGSGHYAAGLELNGTSSGKGEPSSGDHSAGGLVV--------- 51

Qy         61 SEISPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENT 120
              ||:| :||||| || |:| |||||||||||||| ||||||||||||:|| :|||| | :
Db         52 SEVSTISEMPSGSELSTG-DYDYSEEYDNEPQISGYIVDDSVRVEQVIKPKENKTEGEKS 110

Qy        121 SDKPKRKKKGGKNGKNRRNR-KKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGE 179
              |:|||||||| || |||| |||||| |:||||||||||:|||:|| ||| | |:||||
Db        111 SEKPKRKKKGGKGGKGRRNRKKKKNPCAAKFQNFCIHGECRYIENLEVVTCHCHQDYFGE 170

Qy        180 RCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVIT-VQLRRQYVRKYEGEAEER 238
              |||||:|||    || |||||||||  |:||| : |: :|| | ||::: |:|| |||||
Db        171 RCGEKTMKTQKKDDSDLSKIALAAIIVFVSAVSVAAIGIITAVLLRKRFFREYE-EAEER 229

Qy        239 KKLRQENGNVHAIA 252
              ::||||||  ||||
Db        230 RRLRQENGTAHAIA 243


RESULT 7
Q8K3E3
ID   Q8K3E3      PRELIMINARY;      PRT;   165 AA.
AC   Q8K3E3;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Amphiregulin (Fragment).
OS   Mesocricetus auratus (Golden hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Mesocricetus.
OX   NCBI_TaxID=10036;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wang X., Reese J., Paria B.C.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY135447; AAM95451.1; -.
DR   HSSP; Q99075; 1XDT.
FT   NON_TER       1      1
FT   NON_TER     165    165
SQ   SEQUENCE   165 AA;  17986 MW;  7572D02022DAFB26 CRC64;

  Query Match             46.4%;  Score 607.5;  DB 2;  Length 165;
  Best Local Similarity   70.6%;  Pred. No. 3.6e-39;
  Matches  120;  Conservative   16;  Mismatches   25;  Indels    9;  Gaps    4;

Qy          5 LLPPA-PVVLSLLILGSGHYAAGLDLNDTYSGKREPFSGDHSADGFEVTSRSEMSSGSEI 63
              ||| | ||:  |||||||||||||||| | ||| : ||| |:|     |       | |||:
Db          1 LLPRALPVLSLLLILGSGHYAAGLDLNGTTSGKSQAFSGHHNATELVV------SVGSEV 54

Qy         64 SPVSEMPSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDK 123
              | :|||| || |:| |||||||||||||| ||::||||||||:|| :|| | : ||||
Db         55 STISEMPSGSELSTG-DYDYSEEYDNEPQISGYVIDDSVRVEQVIKPKKNKAEGDKTSDK 113

Qy        124 PKRKKKGGKNGKNRRN-RKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKC 172
              |||:|||||:|| ||| :||||||: |||||||||||||:|||| | |
Db        114 PKRRKKGGKSGKGRRNKKKKKNPCDTEFQNFCIHGECKYIENLEAVACNC 163


RESULT 8
Q645M5
ID   Q645M5      PRELIMINARY;      PRT;   210 AA.
AC   Q645M5;
DT   25-OCT-2004 (TrEMBLrel. 28, Created)
DT   25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT   25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE   Amphiregulin.
GN   Name=AREG;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RA   Wang Y., Li J., Leung F.C.;
RT   "Chicken Amphiregulin Gene: cDNA Cloning, Promoter Analysis, and
RT   Regulation of its mRNA Expression in the Chicken Ovary.";
RL   Submitted (AUG-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY725836; AAU21467.1; -.
SQ   SEQUENCE   210 AA;  23665 MW;  B977D786740150CB CRC64;
```

```
  Query Match           33.3%;  Score 435.5;  DB 2;  Length 210;
  Best Local Similarity  48.6%;  Pred. No. 9.1e-26;
  Matches   88;  Conservative   25;  Mismatches   59;  Indels    9;  Gaps    4;

Qy         70 PSSSEPSSGADYDYSEEYDNEPQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKK 129
              |  ||: |:||:  |||:  |    |:| | |||| ||||    |  ||  : ||:|:|
Db         35 PREPEPAPGSDYE-EEEYEEAPLAHQYLVGDLVRVEPVVKPKPAKRGSEKNAGKPRRRKN 93

Qy        130 GGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTH 189
               ||      |:||  ||  |::|||||||| |::||:  |||| |:||||||||: |||
Db         94 KGK------NKKKGTPCEMEYKNFCIHGECVYLQHLQMATCKCYQDYFGERCGEQFMKTQ 147

Qy        190 SMID-SSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVRKYEGEAEERKKLRQENGNV 248
              | :  ||  |  :|  :|::   || :| :   |:   :|| | |||||||||| |
Db        148 RKNDVADYSKTVLVVVAVLLSSISFVAVLIIVIVQVRKKCPQYE-EKEERKKLRQENRNS 206

Qy        249 H 249
              |
Db        207 H 207


RESULT 9
P79199
ID   P79199      PRELIMINARY;      PRT;    77 AA.
AC   P79199;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Amphiregulin (Fragment).
OS   Ovis aries (Sheep).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC   Caprinae; Ovis.
OX   NCBI_TaxID=9940;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Welsh Mountain; TISSUE=Mammary gland;
RX   MEDLINE=97179075; PubMed=9027362; DOI=10.1016/S0303-7207(96)03967-6;
RA   Forsyth I.A., Taylor J.A., Keable S., Turvey A., Lennard S.;
RT   "Expression of amphiregulin in the sheep mammary gland.";
RL   Mol. Cell. Endocrinol. 126:41-48(1997).
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
DR   EMBL; Y09830; CAA70973.1; -.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain.
FT   NON_TER       1      1
FT   NON_TER      77     77
SQ   SEQUENCE   77 AA;  9060 MW;  00BE7B59AE685F25 CRC64;

  Query Match           26.6%;  Score 348;  DB 2;  Length 77;
  Best Local Similarity  77.9%;  Pred. No. 1.5e-19;
  Matches   60;  Conservative    8;  Mismatches    9;  Indels    0;  Gaps    0;

Qy        108 VKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEA 167
              ||| :|||||| |||||||||||||||||||||||:|| |: |||||||||:| ::| ||
Db          1 VKPKKNKTESEKTSDKPKRKKKGGKNGKNRRNRKRKNLCDTEFQNFCIHGKCTFLEQLET 60

Qy        168 VTCKCQQEYFGERCGEK 184
              |:|:|  ||||||||||
Db         61 VSCQCYPEYFGERCGEK 77


RESULT 10
Q6NY79
ID   Q6NY79      PRELIMINARY;      PRT;    239 AA.
AC   Q6NY79;
DT   05-JUL-2004 (TrEMBLrel. 27, Created)
DT   05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT   05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE   LOC407664 protein (Fragment).
GN   Name=LOC407664;
OS   Brachydanio rerio (Zebrafish) (Danio rerio).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC   Cyprinidae; Danio.
OX   NCBI_TaxID=7955;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA   Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA   Jones S.J., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC066705; AAH66705.1; -.
DR   InterPro; IPR001336; EGF_1.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 1.
DR   PRINTS; PR00009; EGFTGF.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 1.
FT   NON_TER       1      1
SQ   SEQUENCE   239 AA;  27024 MW;  49B7FDFDC91925B7 CRC64;

  Query Match           17.2%;  Score 224.5;  DB 2;  Length 239;
  Best Local Similarity  32.4%;  Pred. No. 2.1e-09;
  Matches   68;  Conservative   33;  Mismatches   86;  Indels   23;  Gaps    8;

Qy         45 SADGFEV----TSRSEMSSGSEISPVSEMPSSSEPSSGADYDYSEEY--DNEPQIPGYIV 98
              | | ||      |:   :   || : :: : : :  | | | || |:|  | |  :|
Db         45 SLDRFESGKPRTTEVNLLHTSEKTNITGLQADNGTSVGDD-DYEEDYEGDYELNLP---- 99

Qy         99 DDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGE 158
                 ||     ||       :  |::| |::|: || || | |  ||| |  |:::|||||
Db        100 ----RVAFSTKP--KHPSAMPTTEKTKKRKRKGK-GKGRGN--KKNLCKEEYKDFCIHGV 150

Qy        159 CKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAV 218
              | |   |  :| |  | ||||  ::              ||| ||  :|   |  :||
Db        151 CHYQRDLRTHSCVCHGGYSGERCHVFTLPVGKEEQRYSRTTALAVIAVVLS---LMCLAV 207

Qy        219 ITVQLRRQYVRKYEGEAEERKKLRQENGNV 248
              | : |  :| :| : : |  :|:: |  :|
Db        208 IAILLALRYHKKDDADVESEEKVKLEATSV 237


RESULT 11
```

```
HBGF_CERAE
ID   HBGF_CERAE     STANDARD;      PRT;   208 AA.
AC   Q09118;
DT   01-FEB-1996 (Rel. 33, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Heparin-binding EGF-like growth factor precursor (HB-EGF) (HBEGF)
DE   (Diphteria toxin receptor) (DT-R).
GN   Name=DTR; Synonyms=HEGFL;
OS   Cercopithecus aethiops (Green monkey) (Grivet).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Cercopithecus.
OX   NCBI_TaxID=9534;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92298386; PubMed=1606612;
RA   Naglich J.G., Metherall J.E., Russel D.W., Eidels L.;
RT   "Expression cloning of a diphtheria toxin receptor: identity with a
RT   heparin-binding EGF-like growth factor precursor.";
RL   Cell 69:1051-1061(1992).
RN   [2]
RP   TOXIN-BINDING DOMAIN.
RX   MEDLINE=95126975; PubMed=7826391;
RA   Hooper K.P., Eidels L.;
RT   "Localization of a critical diphtheria toxin-binding domain to the C-
RT   terminus of the mature heparin-binding EGF-like growth factor region
RT   of the diphtheria toxin receptor.";
RL   Biochem. Biophys. Res. Commun. 206:710-717(1995).
RN   [3]
RP   INTERACTION WITH FBLN1.
RX   MEDLINE=22228967; PubMed=11846885; DOI=10.1186/1471-2121-3-2;
RA   Brooke J.S., Cha J.-H., Eidels L.;
RT   "Latent transforming growth factor beta-binding protein-3 and fibulin-
RT   1C interact with the extracellular domain of the heparin-binding EGF-
RT   like growth factor precursor.";
RL   BMC Cell Biol. 3:2-2(2002).
CC   -!- FUNCTION: May be involved in macrophage-mediated cellular
CC       proliferation. It is mitogenic for fibroblasts and smooth muscle
CC       but not endothelial cells. It is able to bind EGF receptors with
CC       higher affinity than EGF itself and is a far more potent mitogen
CC       for smooth muscle cells than EGF (By similarity).
CC   -!- SUBUNIT: Interacts with FBLN1.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Mature HB-EGF is
CC       released into the extracellular space and probably binds to a
CC       receptor (By similarity).
CC   -!- PTM: O-glycosylated (By similarity).
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M93012; -; NOT_ANNOTATED_CDS.
DR   PIR; A41914; A41914.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain; Glycoprotein; Growth factor; Heparin-binding;
KW   Receptor; Signal; Transmembrane.
FT   SIGNAL        1     19       Potential.
FT   PROPEP       20     62       By similarity.
FT   CHAIN        63    148       Heparin-binding EGF-like growth factor.
FT   PROPEP      149    208       C-terminal (Potential).
FT   DOMAIN       20    160       Extracellular (Potential).
FT   TRANSMEM    161    184       Potential.
FT   DOMAIN      185    208       Cytoplasmic (Potential).
FT   DOMAIN      104    144       EGF-like.
FT   CARBOHYD     75     75       O-linked (GalNAc. . .) (By similarity).
FT   CARBOHYD     85     85       O-linked (GalNAc. . .) (By similarity).
FT   DISULFID    108    121       By similarity.
FT   DISULFID    116    132       By similarity.
FT   DISULFID    134    143       By similarity.
SQ   SEQUENCE   208 AA;  22985 MW;  8D108289A0485AE9 CRC64;

  Query Match             16.2%;  Score 211.5;  DB 1;  Length 208;
  Best Local Similarity   36.7%;  Pred. No. 1.7e-08;
  Matches   55;  Conservative   24;  Mismatches   58;  Indels   13;  Gaps    4;

Qy        100 DSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGEC 159
              | :||    ||    | |:   :  ||||||   |      ||::||  ::::|||||||
Db         70 DLLRVTLSSKPQALATPSK--EEHGKRKKKGKGLG------KKRDPCLRKYKDFCIHGEC 121

Qy        160 KYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVI 219
              ||:: | | :| |   | ||||   |:   : : :      || :|  :|:| |    ||
Db        122 KYVKELRAPSCICHPGYHGERCHGLSLPVENRLYTYDHTTILAVVAVVLSSVCL---LVI 178

Qy        220 TVQLRRQYVRK--YEGEAEERKKLRQENGN 247
              |  :| |:  |: | ||: ||   | :
Db        179 VGLLMFRYHRRGGYDVENEEKVKLGMTNSH 208


RESULT 12
HBGF_HUMAN
ID   HBGF_HUMAN     STANDARD;      PRT;   208 AA.
AC   Q99075;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   25-OCT-2004 (Rel. 45, Last annotation update)
DE   Heparin-binding EGF-like growth factor precursor (HB-EGF) (HBEGF)
DE   (Diphteria toxin receptor) (DT-R).
GN   Name=DTR; Synonyms=HEGFL;
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 73-93.
RC   TISSUE=Macrophage;
RX   MEDLINE=91157008; PubMed=1840698;
RA   Higashiyama S., Abraham J.A., Miller J., Fiddes J.C., Klagsbrun M.;
RT   "A heparin-binding growth factor secreted by macrophage-like cells
RT   that is related to EGF.";
RL   Science 251:936-939(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kimmerly W., Bondoc M., Cheng J., Connolly K.S., Gunning K.M.,
RA   Kadner K., Miguel T., Miller C., Pitluck S., Pollard M., Rojeski H.,
RA   Subramanian S., Martin C.H.;
RL   Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Rieder M.J., Livingston R.J., Daniels M.R., Montoya M.A., Chung M.-W.,
RA   Miyamoto K.E., Nguyen C.P., Nguyen D.A., Poel C.L., Robertson P.D.,
RA   Schackwitz W.S., Sherwood J.K., Witrak L.A., Nickerson D.A.;
RT   "NIEHS-SNPs, environmental genome project, NIEHS ES15478, Department
RT   of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).";
RL   Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Eye;
RX   MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
```

```
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [5]
RP   SEQUENCE OF 63-141 AND 143-148, AND CARBOHYDRATE-LINKAGE SITES.
RC   TISSUE=Histiocytic lymphoma;
RX   MEDLINE=92210596; PubMed=1556128;
RA   Higashiyama S., Lau K., Besner G.E., Abraham J.A., Klagsbrun M.;
RT   "Structure of heparin-binding EGF-like growth factor. Multiple forms,
RT   primary structure, and glycosylation of the mature protein.";
RL   J. Biol. Chem. 267:6205-6212(1992).
RN   [6]
RP   TOXIN-BINDING DOMAIN.
RX   MEDLINE=95138082; PubMed=7836353; DOI=10.1074/jbc.270.3.1015;
RA   Mitamura T., Higashiyama S., Taniguchi N., Klagsbrun M., Mekada E.;
RT   "Diphtheria toxin binds to the epidermal growth factor (EGF)-like
RT   domain of human heparin-binding EGF-like growth factor/diphtheria
RT   toxin receptor and inhibits specifically its mitogenic activity.";
RL   J. Biol. Chem. 270:1015-1019(1995).
RN   [7]
RP   X-RAY CRYSTALLOGRAPHY (2.65 ANGSTROMS) OF 73-147 IN COMPLEX WITH TOX.
RX   MEDLINE=98324089; PubMed=9659904; DOI=10.1016/S1097-2765(00)80008-8;
RA   Louie G.V., Yang W., Bowman M.E., Choe S.;
RT   "Crystal structure of the complex of diphtheria toxin with an
RT   extracellular fragment of its receptor.";
RL   Mol. Cell 1:67-78(1997).
CC   -!- FUNCTION: May be involved in macrophage-mediated cellular
CC       proliferation. It is mitogenic for fibroblasts and smooth muscle
CC       but not endothelial cells. It is able to bind EGF receptors with
CC       higher affinity than EGF itself and is a far more potent mitogen
CC       for smooth muscle cells than EGF.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Mature HB-EGF is
CC       released into the extracellular space and probably binds to a
CC       receptor.
CC   -!- PTM: Several N-termini have been identified by direct sequencing.
CC       The forms with N-termini 63, 73 and 74 have been tested and found
CC       to be biologically active.
CC   -!- PTM: O-linked glycan attachment sites were determined by Edman
CC       degradation, O-glycanase digest suggests mucin-type glycosylation
CC       (done in HB-EGF purified from histiocytic lymphoma cell line U-
CC       937).
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M60278; AAA35956.1; -.
DR   EMBL; AC004634; AAC15470.1; -.
DR   EMBL; BC033097; AAH33097.1; -.
DR   EMBL; AY164533; AAN46738.1; -.
DR   PIR; A38432; A38432.
DR   PDB; 1XDT; X-ray; R=69-147.
DR   Genew; HGNC:3059; DTR.
DR   H-InvDB; HIX0005230; -.
DR   MIM; 126150; -.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005887; C:integral to plasma membrane; TAS.
DR   GO; GO:0005154; F:epidermal growth factor receptor binding; TAS.
DR   GO; GO:0007517; P:muscle development; TAS.
DR   GO; GO:0008284; P:positive regulation of cell proliferation; TAS.
DR   GO; GO:0007165; P:signal transduction; TAS.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   3D-structure; Direct protein sequencing; EGF-like domain;
KW   Glycoprotein; Growth factor; Heparin-binding; Receptor; Signal;
KW   Transmembrane.
FT   SIGNAL        1     19       Potential.
FT   PROPEP       20     62       Or 72, or 73, or 76, or 81.
FT   CHAIN        63    148       Heparin-binding EGF-like growth factor.
FT   PROPEP      149    208       C-terminal (Potential).
FT   DOMAIN       20    160       Extracellular (Potential).
FT   TRANSMEM    161    184       Potential.
FT   DOMAIN      185    208       Cytoplasmic (Potential).
FT   DOMAIN      104    144       EGF-like.
FT   DISULFID    108    121       By similarity.
FT   DISULFID    116    132       By similarity.
FT   DISULFID    134    143       By similarity.
FT   CARBOHYD     75     75       O-linked (GalNAc. . .).
FT   CARBOHYD     85     85       O-linked (GalNAc. . .).
FT   TURN        108    114
FT   STRAND      119    121
FT   STRAND      124    124
FT   HELIX       125    127
FT   TURN        128    128
FT   STRAND      129    129
FT   STRAND      132    134
FT   TURN        136    137
FT   STRAND      138    139
FT   TURN        141    142
FT   STRAND      145    146
SQ   SEQUENCE   208 AA;  23067 MW;  2C43C9D1D8291B51 CRC64;

  Query Match             16.2%;  Score 211.5;  DB 1;  Length 208;
  Best Local Similarity   36.7%;  Pred. No. 1.7e-08;
  Matches   55;  Conservative   23;  Mismatches   59;  Indels   13;  Gaps    4;

Qy        100 DSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGEC 159
              | :||    ||    |   |  :  ||||||   |       ||::||  ::::|||||||
Db         70 DLLRVTLSSKPQALAT--PNKEEHGKRKKKGKGLG------KKRDPCLRKYKDFCIHGEC 121

Qy        160 KYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVI 219
              ||:: | | :| |   | ||||   |:   : : :       || :|  :|:| |    ||
Db        122 KYVKELRAPSCICHPGYHGERCHGLSLPVENRLYTYDHTTILAVVAVVLSSVCL---LVI 178

Qy        220 TVQLRRQYVRK--YEGEAEERKKLRQENGN 247
                  :| |:  |: | ||: ||   | :
Db        179 VGLLMFRYHRRGGYDVENEEKVKLGMTNSH 208


RESULT 13
HBGF_PIG
ID   HBGF_PIG       STANDARD;      PRT;   208 AA.
AC   Q01580;
DT   01-FEB-1996 (Rel. 33, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   05-JUL-2004 (Rel. 44, Last annotation update)
DE   Heparin-binding EGF-like growth factor precursor (HB-EGF) (HBEGF).
GN   Name=DTR; Synonyms=HEGFL;
OS   Sus scrofa (Pig).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX   NCBI_TaxID=9823;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Pascall J.C.;
RL   Submitted (NOV-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 99-182 FROM N.A.
RC   STRAIN=Large white; TISSUE=Heart;
RX   MEDLINE=93075016; PubMed=1445231;
RA   Vaughan T.J., Pascall J.C., Brown K.D.;
RT   "Tissue distribution of mRNA for heparin-binding epidermal growth
RT   factor.";
RL   Biochem. J. 287:681-684(1992).
CC   -!- FUNCTION: May be involved in macrophage-mediated cellular
CC       proliferation. It is mitogenic for fibroblasts and smooth muscle
CC       but not endothelial cells. It is able to bind EGF receptors with
CC       higher affinity than EGF itself and is a far more potent mitogen
CC       for smooth muscle cells than EGF (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein. Mature HB-EGF is
CC       released into the extracellular space and probably binds to a
CC       receptor (By similarity).
CC   -!- TISSUE SPECIFICITY: Macrophages, midbrain, cerebellum,
CC       hypothalamus, cerebral cortex, bulbourethral gland, lung, heart
CC       ventricle, kidney, skin, prostate, seminal vesicle, testis; at low
CC       levels in lymph node, thymus, spleen; not detected in pituitary,
CC       olfactory bulb, thyroid, duodenum, pancreas, liver, submaxillary
CC       gland.
CC   -!- PTM: O-glycosylated (Probable).
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Y15731; CAA75740.1; -.
DR   EMBL; X67295; CAA47709.1; -.
DR   PIR; S27162; S27162.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   Pfam; PF00008; EGF; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain; Glycoprotein; Growth factor; Heparin-binding; Signal;
KW   Transmembrane.
FT   SIGNAL       1     23       Potential.
FT   PROPEP       24    62       By similarity.
FT   CHAIN        63    148      Heparin-binding EGF-like growth factor.
FT   PROPEP       149   208      C-terminal (Potential).
FT   DOMAIN       24    161      Extracellular (Potential).
FT   TRANSMEM     162   182      Potential.
FT   DOMAIN       183   208      Cytoplasmic (Potential).
FT   DOMAIN       104   144      EGF-like.
FT   DISULFID     108   121      By similarity.
FT   DISULFID     116   132      By similarity.
FT   DISULFID     134   143      By similarity.
FT   CARBOHYD     85    85       O-linked (GalNAc. . .) (By similarity).
SQ   SEQUENCE   208 AA;  22986 MW;  0A7DA97AE30C8967 CRC64;

  Query Match             15.9%;  Score 208.5;  DB 1;  Length 208;
  Best Local Similarity   32.0%;  Pred. No. 3e-08;
  Matches   62;  Conservative   30;  Mismatches   75;  Indels   27;  Gaps    7;

Qy         54 RSEMSSGSE--ISPVSEMPSSSEPSSGADYDY--SEEYDNEPQIPGYIVDDSVRVEQVVK 109
              |  :: |:   :||        |  |   :    || |          | :| :   |
Db         30 RRGLADGTSNLVSPTESTDQLLPPGGGRGREVLDLEEADL----------DLLRADFSSK 79

Qy        110 PPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFCIHGECKYIEHLEAVT 169
              |    | |:   :: ||||||   |     ||::||  ::::|||||||||:: | | :
Db         80 PQALATPSK--EERGKRKKKGKGLG------KKRDPCLRKYKDFCIHGECKYVKELRAPS 131

Qy        170 CKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSAVILTAVAVITVQLRRQYVR 229
              | |   | ||||   |:   : : :      || :|  :|:| |    ||   |  :| |
Db        132 CICHPGYHGERCHGLSLPVKNRLYTYDHTTILAVVAVVLSSVCL---LVIVGLLMFRYHR 188

Qy        230 K--YEGEAEERKKL 241
              :  |: | ||: ||
Db        189 RGGYDVENEEKVKL 202


RESULT 14
Q9W7C5
ID   Q9W7C5      PRELIMINARY;      PRT;   212 AA.
AC   Q9W7C5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Heparin-binding EGF-like growth factor.
GN   Name=HB-EGF;
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=breed White Leghorn;
RX   MEDLINE=99254107; PubMed=10318950; DOI=10.1073/pnas.96.10.5716;
RA   Fu S.l., Bottoli I., Goller M., Vogt P.K.;
RT   "Heparin-binding epidermal growth factor-like growth factor, a v-Jun
RT   target gene, induces oncogenic transformation.";
RL   Proc. Natl. Acad. Sci. U.S.A. 96:5716-5721(1999).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=breed White Leghorn;
RA   Fu S.L., Bottoli I., Goller M., Vogt P.K.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
DR   EMBL; AF131224; AAD29416.1; -.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain.
SQ   SEQUENCE   212 AA;  22541 MW;  E82A8D08F5297183 CRC64;

  Query Match             15.2%;  Score 198.5;  DB 2;  Length 212;
  Best Local Similarity   33.3%;  Pred. No. 1.8e-07;
  Matches   50;  Conservative   24;  Mismatches   61;  Indels   15;  Gaps    3;

Qy         95 GYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEFQNFC 154
              |  : :  ||  : ||    |        ||:|  | |  | :   ||::||  ::::||
Db         69 GDALSELPRVAFLSKPQGPVT--------PKKKGNGNKRRKGKGLGKKRDPCLRKYKDFC 120

Qy        155 IHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTH---SMIDSSLSKIALAAIAAFMSAV 211
              |||||||  | | :| ||  | ||||    :      |  | : :   :| : : :  |
Db        121 IHGECKYIRELGAPSCICQPGYHGERCHGLLLPVEHPPSTYDHTTALAVVAVVLSSLCLV 180

Qy        212 ILTAVAVITVQLRRQYVRKYEGEAEERKKL 241
              |:||: :     |      |: | ||: ||
Db        181 IITALLMFRCHKR----GVYDVENEEKIKL 206


RESULT 15
Q9BG62
ID   Q9BG62      PRELIMINARY;      PRT;   132 AA.
AC   Q9BG62;
```

```
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Heparin-binding epidermal growth factor (Fragment).
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Klonisch T., Wolf P., Hombach-Klonisch S., Vogt S., Kuechenhoff A.,
RA   Tetens F., Fischer B.;
RL   Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
DR   EMBL; AF333182; AAK14375.1; -.
DR   HSSP; Q99075; 1XDT.
DR   InterPro; IPR001336; EGF_1.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR006210; IEGF.
DR   Pfam; PF00008; EGF; 1.
DR   PRINTS; PR00009; EGFTGF.
DR   SMART; SM00181; EGF; 1.
DR   PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS50026; EGF_3; 1.
KW   EGF-like domain.
FT   NON_TER       1       1
FT   NON_TER     132     132
SQ   SEQUENCE   132 AA;  14628 MW;  6F5ACC07E6A0FF4E CRC64;

  Query Match             14.9%;  Score 195;  DB 2;  Length 132;
  Best Local Similarity   34.9%;  Pred. No. 1.8e-07;
  Matches   44;  Conservative   22;  Mismatches   52;  Indels    8;  Gaps    2;

Qy         91 PQIPGYIVDDSVRVEQVVKPPQNKTESENTSDKPKRKKKGGKNGKNRRNRKKKNPCNAEF 150
              |::|     |  |     ||    | |:   :: ||||||   |       ||::||  ::
Db         15 PEVPDLEEADLYRAAFSSKPQALATPSK--EERGKRKKKGKGLG------KKRDPCLRKY 66

Qy        151 QNFCIHGECKYIEHLEAVTCKCQQEYFGERCGEKSMKTHSMIDSSLSKIALAAIAAFMSA 210
             ::|||||||||:: | | :| |   | ||||   |:   : : :          || :|  :|:
Db         67 KDFCIHGECKYLKELRAPSCICHPGYHGERCHGLSLPVENRLYTYDHTTILAVVAVVLSS 126

Qy        211 VILTAV 216
              |||  :
Db        127 VCLLVI 132


Search completed: February 26, 2005, 16:44:14
Job time : 176 secs
```